

THE DEHN FUNCTION OF $SL(n; \mathbb{Z})$

ROBERT YOUNG

ABSTRACT. We prove that when $n \geq 5$, the Dehn function of $SL(n; \mathbb{Z})$ is quadratic. The proof involves decomposing a disc in $SL(n; \mathbb{R})/SO(n)$ into triangles of varying sizes. By mapping these triangles into $SL(n; \mathbb{Z})$ and replacing large elementary matrices by “shortcuts,” we obtain words of a particular form, and we use combinatorial techniques to fill these loops.

1. INTRODUCTION

The Dehn function is a geometric invariant of a space (typically, a riemannian manifold or a simplicial complex) which measures the difficulty of filling closed curves with discs. This can be made into a group invariant by defining the Dehn function of a group to be the Dehn function of a space on which the group acts cocompactly and properly discontinuously. The choice of space affects the Dehn function, but its rate of growth depends solely on the group. Indeed, it depends solely on the quasi-isometry class of the group.

A small Dehn function can have various geometric implications. In particular, a group with a Dehn function growing slower than n^2 must be hyperbolic; indeed, a group is hyperbolic if and only if its Dehn function grows linearly [Gro87]. The connectedness of the asymptotic cone is also related to the growth of the Dehn function; any asymptotic cone of a group with quadratic Dehn function is simply connected, and conversely, if all asymptotic cones of a group are simply connected, its Dehn function is bounded by a polynomial [Gro93, 5.F'1][Dru98].

One widely-studied family of groups is the set of lattices Γ in semisimple Lie groups G . The Dehn function of a cocompact lattice is easy to find; such a lattice acts on a non-positively curved symmetric space X , and this non-positive curvature gives rise to a linear or quadratic Dehn function, depending on the rank of the space. Non-cocompact lattices have more complicated behavior. The key difference is that if the lattice is not cocompact, it acts cocompactly on a subset of X rather than the whole thing, and this subset has different geometry.

In the case that Γ has \mathbb{Q} -rank 1, the Dehn function is almost completely understood, and depends primarily on the \mathbb{R} -rank of G . In this case, Γ acts cocompactly on a space consisting of X with infinitely many disjoint horoballs removed. When G has \mathbb{R} -rank 1, the boundaries of these horoballs correspond to nilpotent groups, and the lattice is hyperbolic relative to these nilpotent groups. The Dehn function of the lattice is thus equal to that of the nilpotent groups, and Gromov showed that unless X is the complex, quaternionic, or Cayley hyperbolic plane, the Dehn function is at most quadratic [Gro93, 5.A₆]. If X is the complex or quaternionic hyperbolic plane, the Dehn function is cubic [Gro93, Pit97]; if X is the Cayley hyperbolic plane, the precise growth rate is unknown, but is at most cubic.

Date: December 14, 2009.

When G has \mathbb{R} -rank 2 and Γ has \mathbb{Q} -rank 1 or 2, Leuzinger and Pittet [LP96] proved that the Dehn function grows exponentially. As in the \mathbb{R} -rank 1 case, the proof relies on understanding the subgroups corresponding to the removed horoballs, but in this case the subgroups are solvable and have exponential Dehn function. Finally, when G has \mathbb{R} -rank 3 or greater and Γ has \mathbb{Q} -rank 1, Drutu [Dru04] has shown that the boundary of a horoball satisfies a quadratic filling inequality and that Γ enjoys an “asymptotically quadratic” Dehn function, i.e., its Dehn function is bounded by $n^{2+\epsilon}$ for any $\epsilon > 0$.

When Γ has \mathbb{Q} -rank larger than 1, the geometry of the space becomes more complicated. The main difference is that the removed horoballs are no longer disjoint, so many of the previous arguments fail. In many cases, the best known result is due to Gromov, who sketched a proof that the Dehn function of Γ is bounded above by an exponential function [Gro93, 5.A7]. A full proof of this fact was given by Leuzinger [Leu04a].

In this paper, we consider $SL(n; \mathbb{Z})$. This is a lattice with \mathbb{Q} -rank $n-1$ in a group with \mathbb{R} -rank $n-1$, so when n is small, the methods above apply. When $n = 2$, the group $SL(2; \mathbb{Z})$ is virtually free, and thus hyperbolic. As a consequence, its Dehn function is linear. When $n = 3$, the result of Leuzinger and Pittet mentioned above implies that the Dehn function of $SL(3; \mathbb{Z})$ grows exponentially; this was first proved by Epstein and Thurston [ECH⁺92]. This exponential growth has applications to finiteness properties of arithmetic groups as well; Bux and Wortman [BW07] describe a way that the constructions in [ECH⁺92] lead to a proof that $SL(3; \mathbb{F}_q[t])$ is not finitely presented, then generalize to a large class of lattices in reductive groups over function fields.

Much less is known about the Dehn function for lattices in $SL(n; \mathbb{Z})$ when $n \geq 4$. By the results of Gromov and Leuzinger above, the Dehn function of any such lattice is bounded by an exponential function, but Thurston (see [Ger93]) conjectured that

Conjecture 1.1. When $n \geq 4$, $SL(n; \mathbb{Z})$ has a quadratic Dehn function.

This is a special case of a conjecture of Gromov that lattices in symmetric spaces with \mathbb{R} -rank at least 3 have polynomial Dehn functions.

In this paper, we will prove Thurston’s conjecture when $n \geq 5$:

Theorem 1.2. *When $n \geq 5$, $SL(n; \mathbb{Z})$ has a quadratic Dehn function.*

The basic idea of the proof of Theorem 1.2 is to use fillings of curves in the symmetric space $SL(p; \mathbb{R})/SO(p)$ as templates for fillings of words in $SL(p; \mathbb{Z})$. Fillings which lie in the thick part of $SL(p; \mathbb{R})/SO(p)$ correspond directly to fillings in $SL(p; \mathbb{Z})$, but in general, an optimal filling of a curve in the thick part may have to go deep into the cusp of $SL(p; \mathbb{Z}) \backslash SL(p; \mathbb{R})/SO(p)$. Regions of this cusp correspond to parabolic subgroups of $SL(p; \mathbb{Z})$, so our first step is to develop geometric techniques to cut the filling into pieces which each lie in one such region. This reduces the problem of filling the original word to the problem of filling words in parabolic subgroups of Γ . We fill these words using combinatorial techniques, especially the fact that Γ contains many overlapping solvable subgroups.

Some of the ideas in this work were inspired by discussions at the American Institute of Mathematics workshop, “The Isoperimetric Inequality for $SL(n; \mathbb{Z})$,” and the author would like to thank the organizers, Nathan Broaddus, Tim Riley, and Kevin Wortman; and participants, especially Mladen Bestvina, Alex Eskin, Martin Kassabov, and Christophe Pittet. The author would also like to thank

Tim Riley, Yves de Cornulier, Kevin Wortman, and Mladen Bestvina for many helpful conversations while the author was visiting Bristol University, Université de Rennes, and University of Utah. Finally, the author would like to thank New York University for its hospitality during part of the preparation of this paper.

2. PRELIMINARIES

In this section, we will give some preliminaries on the geometry of $SL(n; \mathbb{Z})$.

We use a variant of big-O notation throughout this paper; the notation

$$f(x, y, \dots) = O(g(x, y, \dots))$$

means that there is a $c > 0$ such that $|f(x, y, \dots)| \leq cg(x, y, \dots) + c$ for all values of the parameters.

If $f : X \rightarrow Y$ is Lipschitz, we say that f is c -Lipschitz if $d(f(x), f(y)) \leq cd(x, y)$ for all $x, y \in X$, and we let $\text{Lip}(f)$ be the infimal c such that f is c -Lipschitz.

2.1. Words and curves. If S is a set, we call a formal product of elements of S and their formal inverses a *word* in S . For our purposes, S will usually be a set of generators $\{g_1, \dots\}$ of a group G ; we call a word *reduced* if no generator appears next to its inverse. If

$$u = g_{i(1)}^{\pm 1} g_{i(2)}^{\pm 1} \dots g_{i(n)}^{\pm 1}$$

is a word, we call $\ell_w(u) := n$ its length. We denote the set of words in S by S^* , and if $g, h \in S^*$, we let gh be their concatenation. We denote the empty word by ε . When S is a generating set for G , there is a natural evaluation map from S^* to G , and we say that a word *represents* the corresponding group element.

A word w in S corresponds to a path in the Cayley graph of G with respect to S . This path connects e (the identity element) to the element that w represents, and if w is reduced, the path does not double back on itself. Simplicial paths in the Cayley graph which start at e are in bijective correspondence with words in S .

There is also an approximate version of this correspondence for some group actions on complexes or manifolds. Let X be a simplicial complex or riemannian manifold and let G act on X (by maps of simplicial complexes or by isometries, respectively). We will usually consider either the case that G is a discrete group acting on a complex or the case that $G = X$ is a Lie group or symmetric space. Let $x_0 \in X$. Let $S \subset G$ and for all $g \in S$, let $\gamma_g : [0, 1] \rightarrow G$ be a curve connecting x_0 to gx_0 . If γ_1 and γ_2 are two curves connecting x_0 to g_1x_0 and g_2x_0 respectively, let $\gamma_1\gamma_2$ be the concatenation of γ_1 with the translation $g_1\gamma_2$ of γ_2 . This curve connects x_0 to $g_1g_2x_0$. Similarly, let γ_1^{-1} be the curve connecting x_0 to $g_1^{-1}x_0$ which traces the translation $g_1^{-1}\gamma_1$ in reverse. In this fashion, a word $g_1^{\pm 1} \dots g_n^{\pm 1}$ corresponds to a curve $\gamma_{g_1}^{\pm 1} \dots \gamma_{g_n}^{\pm 1}$ which connects x_0 to $g_1^{\pm 1} \dots g_n^{\pm 1}x_0$. By abuse of notation, we will often use words to denote their corresponding curves. In the next section, we will describe the reverse direction and use the Filling Theorem to approximate curves in G by words in S .

2.2. Dehn functions and the Filling Theorem. The Dehn function is a group invariant providing one way to describe the difficulty of the word problem for the group. The word problem is combinatorial, but the correspondence between words and curves allows us to work with the Dehn function from either a combinatorial or a geometric perspective.

If

$$H = \langle h_1, \dots, h_d \mid r_1, \dots, r_s \rangle$$

is a finitely presented group, we can let $\Sigma = \{h_1, \dots, h_d\}$ and consider words in Σ . If a word w represents the identity, then there is a way to prove this using the relations. That is, there is a sequence of steps which reduces w to the empty word, where each step is a free expansion (insertion of a subword $x_i^{\pm 1} x_i^{\mp 1}$), free reduction (deletion of a subword $x_i^{\pm 1} x_i^{\mp 1}$), or the application of a relator (insertion or deletion of one of the r_i). We call the number of applications of relators in a sequence its *cost*, and we call the minimum cost of a sequence which starts at w and ending at ε the *filling area* of w , denoted by $\delta_H(w)$. We then define the *Dehn function* of H to be

$$\delta_H(n) = \max_{\ell_w(w) \leq n} \delta_H(w),$$

where $\ell_w(w)$ represents the length of w as a word and the maximum is taken over words representing the identity. This depends *a priori* on the chosen presentation of H ; we will see that the growth rate of δ_H is independent of this choice. For convenience, if v, w are two words representing the same element of H , we define $\delta_H(v, w) = \delta_H(vw^{-1})$; this is the minimum cost to transform v to w . This satisfies a triangle inequality in the sense that if w_1, w_2, w_3 are words representing the same element of H , then

$$\delta_H(w_1, w_3) \leq \delta_H(w_1, w_2) + \delta_H(w_2, w_3).$$

This can also be interpreted geometrically. If K_H is the *presentation complex* of H (a simply-connected 2-complex whose 1-skeleton is the Cayley graph of H and whose 2-cells correspond to translates of the relators), then w corresponds to a closed curve in the 1-skeleton of K_H and the sequence of steps reducing w to the identity corresponds to a homotopy contracting this closed curve to a point. More generally, if X is a riemannian manifold or simplicial complex, we can define the filling area $\delta_X(\gamma)$ of a Lipschitz curve $\gamma : S^1 \rightarrow X$ to be the infimal area of a Lipschitz map $D^2 \rightarrow X$ which extends γ . Then we can define the Dehn function of X to be

$$\delta_X(n) = \sup_{\ell_c(\gamma) \leq n} \delta_X(\gamma),$$

where $\ell_c(\gamma)$ is the length of γ as a curve and the supremum is taken over null-homotopic closed curves. As in the combinatorial case, if β and γ are two curves connecting the same points and which are homotopic with their endpoints fixed, we define $\delta_X(\beta, \gamma)$ to be the infimal area of a homotopy between β and γ which fixes their endpoints.

Gromov stated a theorem connecting these two definitions, proofs of which can be found in [Bri02] and [BT02]:

Theorem 2.1 (Gromov's Filling Theorem). *If X is a simply connected riemannian manifold or simplicial complex and H is a finitely presented group acting properly discontinuously, cocompactly, and by isometries on M , then $\delta_H \sim \delta_M$.*

Here, $f \sim g$ if f and g grow at the same rate. Specifically, if $f, g : \mathbb{N} \rightarrow \mathbb{N}$, let $f \lesssim g$ if and only if there is a c such that

$$f(n) \leq cg(cn + c) + c \text{ for all } n$$

and $f \sim g$ if and only if $f \lesssim g$ and $g \lesssim f$. One consequence of Theorem 2.1 is that the Dehn functions corresponding to different presentations of a group are equivalent under this relation.

The idea behind the proof of the Filling Theorem is that under the given conditions, a closed curve in X can be approximated by a word, and a homotopy filling the curve can be approximated by a sequence of applications of relators. Let $\langle R \mid S \rangle$ be a finite presentation for H . We can choose a basepoint and curves in X corresponding to each generator as in the previous section; this corresponds to a choice of an equivariant map from the Cayley graph of H to X . Since X is simply-connected, we can extend this map to a map on K_H . The following lemma, which follows from the Federer-Fleming Deformation Lemma [FF60] or from the Cellulation Lemma [Bri02, 5.2.3], allows us to approximate curves and discs in X by curves (words) and discs in K_H :

Lemma 2.2. *Let H and X be as in the Filling Theorem, and let $f : K_H \rightarrow X$ be an H -equivariant map of a presentation complex for H to X . Then:*

- (1) *Let $s : [0, 1] \rightarrow X$ connect $f(e)$ and $f(h)$, where e is the identity in H and $h \in H$. There is a word w which represents h and which has length $\ell_w(w) \leq c\ell_c(s) + c$. If X is simply connected, then w approximates s in the sense that if $\gamma_w : [0, 1] \rightarrow K_H$ is the curve corresponding to w , then*

$$\delta_X(s, f \circ \gamma_w) = O(\ell_c(s)).$$

- (2) *If w is a word representing the identity in H and $\gamma : S_1 \rightarrow K_H$ is the corresponding closed curve in K_H , then*

$$\delta_H(w) \leq c(\ell_w(w) + \delta_X(f \circ \gamma)).$$

2.3. The geometry of $SL(p; \mathbb{R})$ and $SL(p; \mathbb{Z})$. Let $\Gamma = SL(p; \mathbb{Z})$ and let $G = SL(p) = SL(p; \mathbb{R})$. The group Γ is a lattice in G , and the geometry of G , Γ , and the quotient will be important in our proof. In this section, we will focus on the geometry of G and Γ ; in the next, we will describe the geometry of the quotient.

For our purposes, the main geometric feature of G is that it acts on a non-positively curved symmetric space. Let $\mathcal{E} = SL(p; \mathbb{R})/SO(p)$. The tangent space of \mathcal{E} at the identity, $T_I\mathcal{E}$ is isomorphic to the space of symmetric matrices with trace 0. If u^{tr} represents the transpose of u , then we can define an inner product $\langle u, v \rangle = \text{trace}(u^{tr}v)$ on $T_I\mathcal{E}$. Since this is $SO(p)$ -invariant, it gives rise to a G -invariant riemannian metric on \mathcal{E} . Under this metric, \mathcal{E} is a non-positively curved symmetric space. The lattice Γ acts on \mathcal{E} with finite covolume, but the action is not cocompact. Let $\mathcal{M} := \Gamma \backslash \mathcal{E}$. If $x \in G$, we write the equivalence class of x in \mathcal{E} as $[x]_{\mathcal{E}}$; similarly, if $x \in G$ or $x \in \mathcal{E}$, we write the equivalence class of x in \mathcal{M} as $[x]_{\mathcal{M}}$.

If $g \in G$ is a matrix with coefficients $\{g_{ij}\}$, we define

$$\begin{aligned} \|g\|_2 &= \sqrt{\sum_{i,j} g_{ij}^2}, \\ \|g\|_{\infty} &= \max_{i,j} |g_{ij}|. \end{aligned}$$

Note that for all $g, h \in G$, we have

$$\begin{aligned} \|gh\|_2 &\leq \|g\|_2 \|h\|_2 \\ \|g^{-1}\|_2 &\geq \|g\|_2^{1/p} \end{aligned}$$

and that

$$\log \|g\|_2 = O(d_G(I, g)).$$

The rows of a matrix in G form a unimodular basis of \mathbb{R}^p , so we can interpret G as the space of unimodular bases in \mathbb{R}^p . From this viewpoint, $SO(p)$ acts on a basis by rotating the basis vectors, so \mathcal{E} consists of the set of bases up to rotation. An element of Γ acts by replacing the basis elements by integer combinations of basis elements. This preserves the lattice that they generate, so we can think of $\Gamma \backslash G$ as the set of unit-covolume lattices in \mathbb{R}^p . The quotient $\mathcal{M} = \Gamma \backslash \mathcal{E}$ is then the set of unit-covolume lattices up to rotation. Nearby points in \mathcal{M} or \mathcal{E} correspond to bases or lattices which can be taken into each other by small linear deformations of \mathbb{R}^p . Note that this set is not compact – for instance, the injectivity radius of a lattice is a positive continuous function on \mathcal{M} , and there are lattices with arbitrarily small injectivity radii.

We can use this function to define a subset of \mathcal{E} on which Γ acts cocompactly. Let $\mathcal{E}(\epsilon)$ be the set of points which correspond to lattices with injectivity radius at least ϵ . This is invariant under Γ , and when $0 < \epsilon \leq 1/2$, it is contractible and Γ acts on it cocompactly [ECH⁺92]. We call it the *thick part* of \mathcal{E} , and its preimage $G(\epsilon)$ in G the thick part of G . “Thick” here refers to the fact that the quotients $\Gamma \backslash \mathcal{E}(\epsilon)$ and $\Gamma \backslash G(\epsilon)$ have injectivity radius bounded below. Lemma 2.2 allows us to approximate curves and discs in the thick part of G or \mathcal{E} by words in a finite generating set of Γ and discs in K_Γ , so proving a filling inequality for Γ is equivalent to proving one for $\mathcal{E}(\epsilon)$.

We will also define some subgroups of G . In the following, \mathbb{K} represents either \mathbb{Z} or \mathbb{R} . Let z_1, \dots, z_p be the standard generators for \mathbb{Z}^p , and if $S \subset \{1, \dots, p\}$, let $\mathbb{R}^S = \langle z_s \rangle_{s \in S}$ be a subspace of \mathbb{R}^p . If $q \leq p$, there are many ways to include $SL(q)$ in $SL(p)$. Let $SL(S)$ be the copy of $SL(\#S)$ in $SL(p)$ which acts on \mathbb{R}^S and fixes z_t for $t \notin S$. If S_1, \dots, S_n are disjoint subsets of $\{1, \dots, p\}$ such that $\bigcup S_i = \{1, \dots, p\}$, let

$$U(S_1, \dots, S_n; \mathbb{K}) \subset SL(p; \mathbb{K})$$

be the subgroup of matrices preserving the flag

$$\mathbb{R}^{S_1} \subset \mathbb{R}^{S_1 \cup S_{i-1}} \subset \dots \subset \mathbb{R}^p$$

when acting on the right. If the S_i are sets of consecutive integers in increasing order, $U(S_1, \dots, S_n; \mathbb{K})$ is block upper-triangular. For example, $U(\{1\}, \{2, 3, 4\}; \mathbb{K})$ is the subgroup of $SL(4; \mathbb{K})$ consisting of matrices of the form:

$$\begin{pmatrix} * & * & * & * \\ 0 & * & * & * \\ 0 & * & * & * \\ 0 & * & * & * \end{pmatrix}.$$

If $d_1, \dots, d_n > 0$ and $\sum_i d_i = p$, let $U(d_1, \dots, d_n; \mathbb{K})$ be the group of upper block triangular matrices with blocks of the given lengths, so that the subgroup illustrated above is $U(1, 3; \mathbb{K})$. Each group $U(d_1, \dots, d_n; \mathbb{Z})$ is a parabolic subgroup of Γ , and any parabolic subgroup of Γ is conjugate to a unique such group. Let \mathcal{P} be the set of groups $U(d_1, \dots, d_n; \mathbb{Z})$, including $U(p; \mathbb{Z}) = \Gamma$.

One feature of $SL(p; \mathbb{Z})$ is its particularly simple presentation. If $1 \leq i \neq j \leq p$, let $e_{ij}(x) \in \Gamma$ be the matrix which consists of the identity matrix with the (i, j) -entry replaced by x ; we call these *elementary matrices*. Let $e_{ij} := e_{ij}(1)$. When

$p \geq 3$, there is a finite presentation which has the matrices e_{ij} as generators [Mil71]:

$$(1) \quad \begin{aligned} \Gamma = \langle e_{ij} \mid & [e_{ij}, e_{kl}] = I && \text{if } i \neq l \text{ and } j \neq k \\ & [e_{ij}, e_{jk}] = e_{ik} && \text{if } i \neq k \\ & (e_{ij} e_{ji}^{-1} e_{ij})^4 = I \rangle, \end{aligned}$$

where we adopt the convention that $[x, y] = xyx^{-1}y^{-1}$. We will use a slightly expanded set of generators. Let

$$\Sigma := \{e_{ij} \mid 1 \leq i \neq j \leq p\} \cup D,$$

where $D \subset \Gamma$ is the set of diagonal matrices in $SL(p; \mathbb{Z})$; note that this set is finite. If R is the set of relators given above with additional relations expressing each element of D as a product of elementary matrices, then $\langle \Sigma \mid R \rangle$ is a finite presentation of Γ with relations R . Furthermore, if $H = SL(q; \mathbb{Z}) \subset SL(p; \mathbb{Z})$ or if H is a subgroup of block-upper-triangular matrices, then H is generated by $\Sigma \cap H$.

For each $s \in \Sigma$, associate s with a geodesic in $\mathcal{E}(1/2)$ connecting $[I]_{\mathcal{E}}$ to $[s]_{\mathcal{E}}$. Let c_{Σ} be the maximum length of one of these curves. These give a correspondence between words in Σ and curves in \mathcal{E} ; by abuse of notation, we will often use words in Σ to refer to their corresponding curves in \mathcal{E} .

One key fact about the geometry of $SL(p; \mathbb{Z})$ is a theorem of Lubotzky, Mozes, and Raghunathan [LMR93]:

Theorem 2.3. *The word metric on $SL(p; \mathbb{Z})$ for $p \geq 3$ is equivalent to the restriction of the Riemannian metric of $SL(p; \mathbb{R})$ to $SL(p; \mathbb{Z})$. That is, there is a c such that for all $g \in SL(p; \mathbb{Z})$, we have*

$$c^{-1}d_G(I, g) \leq d_{\Gamma}(I, g) \leq cd_G(I, g).$$

Part of the proof of this theorem is a construction of short words which represent large elementary matrices. Lubotzky, Mozes, and Raghunathan construct these words by including the solvable group $\mathbb{R} \ltimes \mathbb{R}^2$ in the thick part of G . Since $\mathbb{R}^2 \subset \mathbb{R} \ltimes \mathbb{R}^2$ is exponentially distorted, there are short curves in $\mathbb{R} \ltimes \mathbb{R}^2$ connecting I to $e_{ij}(x)$ which can be approximated by short words in Γ . For our purposes, we will need a more general construction.

Lemma 2.4. *Let $p \geq 3$, and let $S, T \subset \{1, \dots, p\}$ be disjoint. Let $s = \#S$ and $t = \#T$, and assume that $s \geq 2$. If $V \in \mathbb{R}^S \otimes \mathbb{R}^T$, define*

$$u(V) = u(V; S, T) := \begin{pmatrix} I_S & V \\ 0 & I_T \end{pmatrix} \in U(S, T).$$

There is a subgroup

$$H_{S,T} \cong (\mathbb{R}^{s-1} \times \mathbb{R}^{t-1}) \ltimes (\mathbb{R}^S \otimes \mathbb{R}^T) \subset U(S, T)$$

which lies in $G(\epsilon)$ and a family of curves $\hat{u}(V; S, T) : [0, 1] \rightarrow H_{S,T}$ connecting I to $u(V; S, T)$ such that $\ell_c(\hat{u}(V; S, T)) = O(\log \|V\|_2)$, where $\log x = \max\{1, \log x\}$.

Proof. First, we define $H_{S,T} \subset U(S, T)$. Let A_1, \dots, A_s be a set of simultaneously diagonalizable positive-definite matrices in $SL(S; \mathbb{Z})$. The A_i 's have the same eigenvectors; call these shared eigenvectors $v_1, \dots, v_s \in \mathbb{R}^S$, and normalize them to have unit length. The A_i are entirely determined by their eigenvalues, and we can define vectors

$$q_i = (\log \|A_i v_1\|_2, \dots, \log \|A_i v_s\|_2) \in \mathbb{R}^s$$

Since $A_i \in SL(S; \mathbb{Z})$, the product of its eigenvalues is 1, and the sum of the coordinates of q_i is 0. We require that the A_i are independent in the sense that the q_i span a $(s-1)$ -dimensional subspace of \mathbb{R}^s ; since they are all contained in an $(s-1)$ -dimensional subspace, this is the maximum rank possible. If a set of matrices satisfies these conditions, we call them a set of *independent commuting matrices* for S . A construction of such matrices can be found in Section 10.4 of [ECH⁺92]. The A_i generate a subgroup isomorphic to \mathbb{Z}^{s-1} , and by possibly choosing a different generating set for this subgroup, we can assume that $\lambda_i := \|A_i v_i\|_2 > 1$ for all i .

If $t \geq 2$, let $B_1^{tr}, \dots, B_t^{tr} \in SL(T; \mathbb{Z})$ (where tr represents the transpose of a matrix) be a set of independent commuting matrices for T and let $w_1, \dots, w_t \in \mathbb{R}^T$ be a basis of unit eigenvectors of the B_i^{tr} . If $t = 1$, let $B_1 = (1)$ and let w_1 be a generator of $\mathbb{R}^T = \mathbb{R}$. Let

$$H_{S,T} := \left\{ \begin{pmatrix} \prod_i A_i^{x_i} & V \\ 0 & \prod_i B_i^{y_i} \end{pmatrix} \middle| x_i, y_i \in \mathbb{R}, V \in \mathbb{R}^S \otimes \mathbb{R}^T \right\} \\ = (\mathbb{R}^{s-1} \times \mathbb{R}^{t-1}) \ltimes (\mathbb{R}^S \otimes \mathbb{R}^T).$$

When the x_i and y_i are integers and V has integer coefficients, this matrix has integer coefficients. Furthermore, $H_{S,T} \cap \Gamma$ is a cocompact lattice in $H_{S,T}$, so $H_{S,T}$ is contained in $G(\epsilon)$ for some $\epsilon > 0$.

By abuse of notation, let A_i and B_i refer to the corresponding matrices in $H_{S,T}$. The group $H_{S,T}$ is generated by powers of the A_i , powers of the B_i , and elementary matrices in the sense that any element of $H_{S,T}$ can be written as

$$\prod A_i^{x_i} \prod B_i^{y_i} \begin{pmatrix} I_S & V \\ 0 & I_T \end{pmatrix},$$

for some $x_i, y_i \in \mathbb{R}$ and $V \in \mathbb{R}^S \otimes \mathbb{R}^T$, where I_S and I_T represent the identity matrix in $SL(S; \mathbb{Z})$ and $SL(T; \mathbb{Z})$ respectively. Let

$$W = \{A_i^x\}_{x \in \mathbb{R}} \cup \{B_i^x\}_{x \in \mathbb{R}} \cup \{u(V)\}_{V \in \mathbb{R}^S \otimes \mathbb{R}^T}.$$

We let A_i^x correspond to the curve

$$d \mapsto \begin{pmatrix} A_i^{xd} & 0 \\ 0 & I_T \end{pmatrix},$$

B_i^x to the curve

$$d \mapsto \begin{pmatrix} I_S & 0 \\ 0 & B_i^{xd} \end{pmatrix},$$

and

$$u(V) = \begin{pmatrix} I_S & V \\ 0 & I_T \end{pmatrix}$$

to the curve

$$d \mapsto \begin{pmatrix} I_S & dV \\ 0 & I_T \end{pmatrix},$$

where in all cases, d ranges from 0 to 1. This gives a correspondence between elements of W and curves. Let $c \geq \max_i \{\ell_c(A_i)\}$. The word $A_i^x u(v_i \otimes w) A_i^{-x}$ represents the matrix $u(\lambda_i^x v_i \otimes w)$ and corresponds to a curve of length at most $2cx + \|v_i\|_2 \|w\|_2$ connecting I and $u(\lambda_i^x v_i \otimes w)$.

If $V \in \mathbb{R}^S \otimes \mathbb{R}^T$, then

$$V = \sum_{i,j} x_{ij} v_i \otimes w_j$$

for some $x_{ij} \in \mathbb{R}$. Let

$$\gamma_{ij}(x) = \begin{cases} A_i^{\log_{\lambda_i} |x|} u(\text{sign}(x) v_i \otimes w_j) A_i^{-\log_{\lambda_i} |x|} & \text{if } |x| > 1, \\ u(x v_i \otimes w_j) & \text{if } |x| \leq 1, \end{cases}$$

where $\text{sign}(x) = \pm 1$ depending on whether x is positive or negative. Let

$$\widehat{u}(V) := \prod_{i,j} \gamma_{ij}(x_{ij}).$$

Then $\widehat{u}(V)$ represents $u(V)$ and

$$\ell_c(\widehat{u}(V)) = O(\overline{\log} \|V\|_2)$$

for all V , where $\overline{\log} x = \max\{1, \log x\}$. \square

If $i \in S$ and $j \in T$, then $e_{ij}(x)$ and $u(x z_i \otimes z_j; S, T)$ represent the same matrix. If $x \in \mathbb{Z}$, then we define $\widehat{e}_{ij;S,T}(x) \in \Sigma^*$ to be a word approximating $\widehat{u}(x z_i \otimes z_j; S, T)$. Changing S and T changes $\widehat{e}_{ij;S,T}(x)$, but in Sec. 4.2, we will prove that in many cases, $\widehat{e}_{ij;S,T}(x)$ and $\widehat{e}_{ij;S',T'}(x)$ are connected by a homotopy of area $O((\log |x|)^2)$. Because of this, the choice of S and T is largely irrelevant. Thus, for each (i, j) , we choose a $d \notin \{i, j\}$ and let

$$\widehat{e}_{ij}(x) = \widehat{e}_{ij;\{i,d\},\{j\}}(x).$$

These “shortcuts” are key to our proof. Just as Lubotzky, Mozes, and Raghunathan build paths in Γ out of the $\widehat{e}_{ij}(x)$, we will build fillings in Γ out of fillings of words involving the $\widehat{e}_{ij}(x)$. With this in mind, we introduce an infinite generating set for Γ . Let

$$\widehat{\Sigma} := \{e_{ab}(x) \mid x \in \mathbb{Z} - \{0\}, a, b \in \{1, \dots, p\}, a \neq b\} \cup D.$$

This set contains Σ , so it generates Γ . Define a map $\lambda : \widehat{\Sigma}^* \rightarrow \Sigma^*$ which sends elements of Σ to themselves and sends $e_{ab}(x)$ to $\widehat{e}_{ab}(x)$, and define a length function $\widehat{\ell} : \widehat{\Sigma}^* \rightarrow \mathbb{Z}$ by

$$\widehat{\ell}(w) = \ell_w(\lambda(w)).$$

Words in $\widehat{\Sigma}$ are associated to curves in the thick part of G , and in Section 4.2, we will construct discs filling several families of such curves. We will use these “relations” to manipulate words in $\widehat{\Sigma}$.

2.4. The geometry of \mathcal{M} . Since \mathcal{E} is non-positively curved, $\delta_{\mathcal{E}}$ grows at most quadratically, and our goal is to show that the same is true for the subset $\mathcal{E}(\epsilon)$. In order to do so, we must study the *thin part* $\mathcal{E} - \mathcal{E}(\epsilon)$ of \mathcal{E} , and since the cusp of \mathcal{M} corresponds to this thin part, we will study \mathcal{M} . The constructions in this section generalize to many reductive and semisimple Lie groups with the use of precise reduction theory, but we will only state the results for $SL(p; \mathbb{Z})$, as stating the theorems in full generality requires a lot of additional background.

Let $\text{diag}(t_1, \dots, t_p)$ be the diagonal matrix with entries (t_1, \dots, t_p) . Let A be the set of diagonal matrices in G and if $\epsilon > 0$, let

$$A_{\epsilon}^+ = \{\text{diag}(t_1, \dots, t_p) \mid \prod t_i = 1, t_i > 0, t_i \geq \epsilon t_{i+1}\}.$$

One of the main features of \mathcal{M} is that it is Hausdorff equivalent to A_{ϵ}^+ ; our main goal in this section is to describe this Hausdorff equivalence and its “fibers”. Let N be the set of upper triangular matrices with 1’s on the diagonal and let N^+ be

the subset of N with off-diagonal entries in the interval $[-1/2, 1/2]$. Translates of the set $N^+A_\epsilon^+$ are known as Siegel sets. The following properties of Siegel sets are well known (see for instance [BHC62]).

Lemma 2.5.

There is an $1 > \epsilon_S > 0$ such that if we let

$$\mathcal{S} := [N^+A_{\epsilon_S}^+]_{\mathcal{E}} \subset \mathcal{E},$$

then

- $\Gamma\mathcal{S} = \mathcal{E}$.
- There are only finitely many elements $\gamma \in \Gamma$ such that $\gamma\mathcal{S} \cap \mathcal{S} \neq \emptyset$.

In particular, the quotient map $\mathcal{S} \rightarrow \mathcal{M}$ is a surjection. We define $A^+ := A_{\epsilon_S}^+$.

The inclusion $A^+ \hookrightarrow \mathcal{S}$ is a Hausdorff equivalence:

Lemma 2.6 ([JM02]). *Give A the riemannian metric inherited from its inclusion in G , so that*

$$d_A(\text{diag}(d_1, \dots, d_p), \text{diag}(d'_1, \dots, d'_p)) = \sqrt{\sum_{i=1}^p \left| \log \frac{d'_i}{d_i} \right|^2}.$$

- There is a c such that if $n \in N^+$ and $a \in A^+$, then $d_{\mathcal{E}}([na]_{\mathcal{E}}, [a]_{\mathcal{E}}) \leq c$. In particular, if $x \in \mathcal{S}$, then $d_{\mathcal{E}}(x, [A^+]_{\mathcal{E}}) \leq c$.
- If $x, y \in A^+$, then $d_A(x, y) = d_{\mathcal{S}}(x, y)$.

Proof. For the first claim, note that if $x = [na]_{\mathcal{E}}$, then $x = [a(a^{-1}na)]_{\mathcal{E}}$, and $a^{-1}na \in N$. Furthermore,

$$\|a^{-1}na\|_{\infty} \leq \epsilon_S^p,$$

so

$$d_{\mathcal{E}}([x]_{\mathcal{E}}, [a]_{\mathcal{E}}) \leq d_G(I, a^{-1}na)$$

is bounded independently of x .

For the second claim, we clearly have $d_A(x, y) \geq d_{\mathcal{S}}(x, y)$. For the reverse inequality, it suffices to note that the map $\mathcal{S} \rightarrow A^+$ given by $na \mapsto a$ for all $n \in N^+$, $a \in A^+$ is distance-decreasing. \square

Siegel conjectured that the quotient map from \mathcal{S} to \mathcal{M} is also a Hausdorff equivalence, that is:

Theorem 2.7. *There is a c' such that if $x, y \in \mathcal{S}$, then*

$$d_{\mathcal{E}}(x, y) - c' \leq d_{\mathcal{M}}([x]_{\mathcal{M}}, [y]_{\mathcal{M}}) \leq d_{\mathcal{E}}(x, y)$$

Proofs of this conjecture can be found in [Leu04b, Ji98, Din94]. As a consequence, the natural quotient map $A^+ \rightarrow \mathcal{M}$ is a Hausdorff equivalence.

Any point $x \in \mathcal{E}$ can be written (possibly non-uniquely) as $x = [\gamma na]_{\mathcal{E}}$ for some $\gamma \in \Gamma$, $n \in N^+$ and $a \in A^+$. The theorem implies that these different decompositions are a bounded distance apart:

Corollary 2.8 (see [JM02], Lemmas 5.13, 5.14). *There is a constant c'' such that if $x, y \in \mathcal{M}$, $n, n' \in N^+$ and $a, a' \in A^+$ are such that $x = [na]_{\mathcal{M}}$ and $y = [n'a']_{\mathcal{M}}$, then*

$$|d_{\mathcal{M}}(x, y) - d_A(a, a')| \leq c''.$$

In particular, if $[na]_{\mathcal{M}} = [n'a']_{\mathcal{M}}$, then $d_A(a, a') \leq c''$.

Proof. Note that by Lemma 2.6,

$$d_{\mathcal{M}}(x, [a]_{\mathcal{M}}) \leq d_{\mathcal{E}}([na]_{\mathcal{E}}, [a]_{\mathcal{E}}) \leq c$$

and likewise $d_{\mathcal{M}}(y, [a']_{\mathcal{M}}) \leq c$. Furthermore, by the theorem and the lemma,

$$d_A(a, a') - c' = d_S(a, a') - c' \leq d_{\mathcal{M}}([a]_{\mathcal{M}}, [a']_{\mathcal{M}}) \leq d_A(a, a'),$$

so if we let $c'' = c' + 2c$, the corollary follows. \square

Let $\rho : \mathcal{E} \rightarrow \Gamma$ be a map such that $\rho(\mathcal{S}) = I$ and $x \in \rho(x)\mathcal{S}$ for all x . Any point $x \in \mathcal{E}$ can be uniquely written as $x = [\rho(x)na]_{\mathcal{E}}$ for some $n \in N^+$ and $a \in A^+$. Let $\phi : \mathcal{E} \rightarrow A^+$ be the map $[\rho(x)na]_{\mathcal{E}} \mapsto a$. If $\phi(x) = \text{diag}(a_1, \dots, a_p)$, let $\phi_i(x) = \log a_i$. If $x, y \in \mathcal{E}$, then $|\phi_i(x) - \phi_i(y)| \leq d_{\mathcal{E}}(x, y) + c''$; let $c_{\phi} := c''$.

If $x \in \mathcal{E}$, then the values of ρ near x are largely determined by $\phi(x)$. For instance, if $\|\phi(x)\|_2$ is small, then x is in the thick part of \mathcal{E} , and so if y is close to x , then $\rho(x)$ is close to $\rho(y)$. If $\|\phi(x)\|_2$ is large, then $[x]_{\mathcal{M}}$ is deep in the cusp of \mathcal{M} and so $\rho(y)$ may be exponentially far from $\rho(x)$. Parts of the cusp, however, generally have simpler topology than the entire quotient.

Recall that if $\tilde{x} \in G$ is a representative of x , we can construct a lattice $\mathbb{Z}^p \tilde{x} \subset \mathbb{R}^p$. If $\|\phi(x)\|_2$ is large, then $\mathbb{Z}^p \tilde{x}$ has short vectors. In the following lemmas, we will use the subspaces generated by these short vectors to show that the fundamental group of a small ball in the cusp is contained in a parabolic subgroup. Recall that $z_1, \dots, z_p \in \mathbb{Z}^p$ are the standard generators of \mathbb{Z}^p .

Lemma 2.9. *Let $x = [\gamma na]_{\mathcal{E}}$ for some $\gamma \in \Gamma$, $n \in N^+$, and $a \in A^+$. Let $\tilde{x} \in G$ be such that $x = [\tilde{x}]_{\mathcal{E}}$ and let*

$$V(x, r) = \langle v \in \mathbb{Z}^p \mid \|v\tilde{x}\|_2 \leq r \rangle.$$

Different choices of \tilde{x} differ by an element of $SO(p)$, so $V(x, r)$ is independent of the choice of \tilde{x} . We claim there is a $c_V > 0$ such that if $a = \text{diag}(a_1, \dots, a_p)$ and

$$e^{c_V} a_{k+1} < r < e^{-c_V} a_k,$$

then $V(x, r) = Z_k \gamma^{-1}$, where $Z_k := \langle z_{k+1}, \dots, z_p \rangle$.

Proof. Note that $V(\gamma' x', r) = V(x', r) \gamma'^{-1}$, so we may assume that $\gamma = I$ without loss of generality. Let $n = \{n_{ij}\} \in N^+$ and let $\tilde{x} = na$. We have

$$\begin{aligned} z_j \tilde{x} &= z_j na \\ &= a_j z_j + \sum_{i=j+1}^p n_{ji} z_i a_i. \end{aligned}$$

Since $a_{i+1} \leq a_i \epsilon_S^{-1}$, we have $a_i \leq a_{k+1} \epsilon_S^{-p}$ for $i \geq k+1$ and $a_i \geq a_k \epsilon_S^p$ for $i \leq k$. Since $|n_{ji}| \leq 1/2$ when $i > j$, we have

$$\|z_j \tilde{x}\|_2 \leq a_{k+1} \sqrt{p} \epsilon_S^{-p}$$

when $j > k$. Thus

$$V(x, a_{k+1} \sqrt{p} \epsilon_S^{-p}) \supset Z_k.$$

On the other hand, assume that $v \notin Z_k$, and let $v = \sum_i v_i z_i$ for some $v_i \in \mathbb{Z}$. Let j be the smallest integer such that $v_j \neq 0$; by assumption, $j \leq k$. The z_j -coordinate of $v\tilde{x}$ is $v_j a_j$, so

$$\|v\tilde{x}\|_2 \geq a_j > a_k \epsilon_S^p$$

and thus if $t < a_k \epsilon_S^p$, then $V(x, t) \subset Z_k$. Therefore, if

$$a_{k+1} \sqrt{p} \epsilon_S^{-p} \leq t < a_k \epsilon_S^p,$$

then $V(\tilde{x}, t) = Z_j$. We can choose $c_V = \log \sqrt{p} \epsilon_S^{-p}$. \square

Lemma 2.10. *There is a constant c such that if $1 \leq j < p$ and $x, y \in \mathcal{E}$ are such that*

$$d_{\mathcal{E}}(x, y) < \frac{\phi_j(x) - \phi_{j+1}(x)}{2} - c$$

and $g, h \in \Gamma$ are such that $x \in g\mathcal{S}$ and $y \in h\mathcal{S}$, then $g^{-1}h \in U(j, p-j)$.

Proof. Let

$$r = \exp \frac{\phi_j(x) + \phi_{j+1}(x)}{2},$$

and let $c = c_\phi + c_V$, so that we have

$$\phi_{j+1}(x) + c + d_{\mathcal{E}}(x, y) < \log r < \phi_j(x) - c - d_{\mathcal{E}}(x, y).$$

By the previous lemma,

$$V(x, r \exp(-d_{\mathcal{E}}(x, y))) = V(x, r \exp(d_{\mathcal{E}}(x, y))) = Z_j g^{-1}.$$

Furthermore, $|\phi_i(x) - \phi_i(y)| < d_{\mathcal{E}}(x, y) + c_\phi$, so

$$\phi_{j+1}(y) - c_\phi + c < \log r < \phi_j(y) + c_\phi - c$$

and $V(y, r) = Z_j h^{-1}$.

Note that since

$$\|v\tilde{y}\|_2 \exp(-d_{\mathcal{E}}(x, y)) \leq \|v\tilde{x}\|_2 \leq \|v\tilde{y}\|_2 \exp(d_{\mathcal{E}}(x, y)),$$

we have

$$V(x, r \exp(-d_{\mathcal{E}}(x, y))) \subset V(y, r) \subset V(x, r \exp(d_{\mathcal{E}}(x, y))),$$

so $Z_j g^{-1}h = Z_j$. Since $g^{-1}h$ stabilizes Z_j , we have $g^{-1}h \in U(j, p-j)$ as desired. \square

The differences $\phi_j(x) - \phi_{j+1}(x)$ increase as the distance between x and Γ increases:

Corollary 2.11. *There is a $c' > 0$ such that if*

$$d_{\mathcal{E}}(x, y) < \frac{d_{\mathcal{M}}([x]_{\mathcal{M}}, [I]_{\mathcal{M}})}{2p^2} - c',$$

and $g, h \in \Gamma$ are such that $x \in g\mathcal{S}$ and $y \in h\mathcal{S}$, then there is a j depending only on x such that $g^{-1}h \in U(j, p-j)$.

Proof. We will find a c' such that if the hypothesis is satisfied, then

$$d_{\mathcal{E}}(x, y) < \frac{\phi_j(x) - \phi_{j+1}(x)}{2} - c$$

for some j , where c is as in Lemma 2.10.

Since $\sum_i \phi_i(x) = 0$, we have

$$|\phi_j(x)| \leq p \max_i |\phi_i(x) - \phi_{i+1}(x)|,$$

and since $\phi_i(x) - \phi_{i+1}(x)$ is bounded below, there is a c_0 such that

$$\frac{d_A(I, \phi(x))}{p^2} \leq \max_i (\phi_i(x) - \phi_{i+1}(x)) + c_0.$$

Furthermore, we have

$$d_A(I, \phi(x)) \geq d_{\mathcal{M}}([x]_{\mathcal{M}}, [I]_{\mathcal{M}}) - c_{\phi}$$

by the definition of c_{ϕ} .

Combining these two inequalities, we find that there is a $c' > 0$ such that

$$d_{\mathcal{E}}(x, y) < \frac{\phi_j(x) - \phi_{j+1}(x)}{2} - c$$

for some j , so x and y satisfy the conditions of Lemma 2.10. \square

2.5. Sketch of proof. The basic idea behind the proof of Theorem 1.2 is to use a filling in \mathcal{E} as a template for a filling in Γ . Since \mathcal{E} is non-positively curved, it has a quadratic Dehn function, so a word in Γ can be filled by a quadratic disc in \mathcal{E} . The problem with this filling is that it is likely to leave the thick part of \mathcal{E} and go deep into the cusp of \mathcal{M} , so we need to replace the parts of the filling that lie in the cusp with pieces in the thick part of \mathcal{E} .

To do this, we use Cor. 2.11, which implies that piece of the filling which lie in the cusp actually lie in parabolic subgroup. This lets us fill them using an inductive argument.

The argument breaks down into two main pieces. The first is primarily geometric: we partition \mathcal{M} into pieces corresponding to parabolic subgroups and break a filling in \mathcal{E} into triangles which each lie in one of the pieces. The second is more combinatorial: we use combinatorial methods to fill each of these triangles.

3. THE GEOMETRIC STEP: CREATING A TEMPLATE

As curves in a group grow longer, they can increase in both size and complexity. Many bounds on Dehn functions come from a combination of techniques: one to reduce a complex curve to simple curves, and one to fill those simple curves. This section is devoted to the former problem; we will use the geometry of \mathcal{M} to reduce a curve in Γ to a collection of simpler curves in parabolic subgroups.

We will first describe a framework for reducing curves to triangles which appears frequently in proofs of filling inequalities. Let H be a group generated by S and let $\omega : H \rightarrow S^*$ be a map such that $\omega(h)$ represents h for all $h \in H$ and $\ell_w(\omega(h)) = O(d(I, h))$. We call this a normal form for H ; note that it does not have to satisfy a fellow-traveler condition. Let τ be a directed planar graph whose boundary is a circle, whose internal faces have either two or three sides, and whose vertices are labeled by elements of h . We think of τ as a triangulation of D^2 . Typically, to fill a word $w = w_1 \dots w_n$, we use a τ whose boundary is an n -gon with labels $I, w_1, w_1 w_2, \dots, w_1 \dots w_{n-1}$; we call this a *template for w* .

We can use τ and ω to construct a map $D^2 \rightarrow K_H$. First, we send each vertex to its label. Second, if e is an edge from a vertex labeled h_1 to a vertex labeled h_2 , we send e to $h_1 \cdot \omega(h_1^{-1} h_2)$, which connects h_1 and h_2 . This sends the boundary of τ to a word $\omega(h_1)^{\pm 1} \dots \omega(h_n)^{\pm 1}$ which we call the *boundary word* of τ . If τ is a template for w and w_{τ} is its boundary word, then $\delta(w, w_{\tau}) \leq c \ell_w(w)$, for some c depending only on H and ω . It sends the boundary of each triangle to a product

$$\omega(h_1)^{\pm 1} \omega(h_2)^{\pm 1} \omega(h_3)^{\pm 1},$$

which we call an ω -triangle, and likewise, each bigon to a product

$$\omega(h_1)^{\pm 1} \omega(h_2)^{\pm 1}.$$

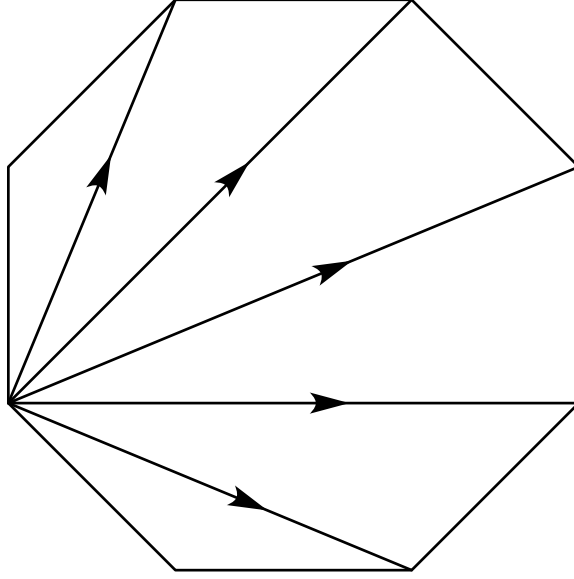


FIGURE 1. A seashell template

We think of a bigon as a degenerate triangle in which $h_3 = I$. By filling the ω -triangles and bigons, we get a filling of the boundary word of τ whose area depends on τ and on the difficulty of filling ω -triangles and bigons.

In many cases, a good choice of ω and of τ makes these ω -triangles easy to fill. One example is the case of automatic groups; in an automatic group, ω is the automatic structure, and τ is the “seashell” template in Fig. 1, whose vertices are labeled $I, w_1, w_1w_2, \dots, w_1 \dots w_{n-1}$. Each triangle then has two edges in normal form and one short edge, and the fellow-traveler property of ω lets us fill each such triangle with a disc of area $O(n)$. Since there are n triangles, this gives a quadratic bound on the Dehn function. Similarly, some proofs (for example, [GHR03]) define a normal form ω for elements of H and describe homotopies between $\omega(h)s$ and $\omega(hs)$ for all $h \in H$ and $s \in S$; these can also be described in terms of a seashell template.

Another application is the following proposition, which can be proved using templates like that in Fig. 2. It implies that if ω -triangles can be filled by discs of polynomial area, then so can arbitrary curves.

Proposition 3.1. *If there is an $\alpha > 1$ such that for all $h_i \in H$ such that h_1h_2 and $d(I, h_i) \leq \ell$, we have*

$$\delta_H(\omega(h_1)\omega(h_2)\omega(h_1h_2)^{-1}) = O(\ell^\alpha),$$

then $\delta_H(n) \lesssim n^\alpha$.

Proof. Without loss of generality, we may assume that the identity I is in the generating set of H . It suffices to consider the case that $\ell_w(w) = n = 2^k$ for some $k \in \mathbb{Z}$; otherwise, we may pad w with the letter I until its length is a power of 2. Let τ be the template consisting of $2^k - 2$ triangles and 1 bigon as in Fig. 2.

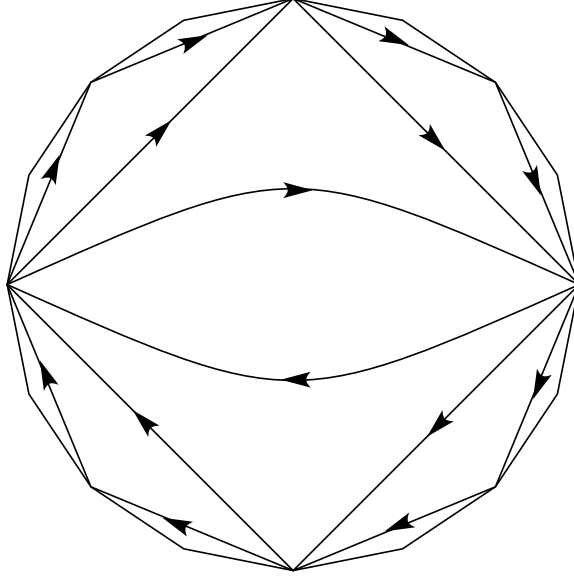


FIGURE 2. A dyadic template

Let $w = w_1 \dots w_n$ and let $w(i) = w_1 \dots w_i$. Label the vertices of the template by $w(i)$. Each triangle has vertices labeled

$$w(i2^j), w((i + \frac{1}{2})2^{j-1}), w((i+1)2^j)$$

for some $1 \leq j < k$ and $0 \leq i < 2^{k-j}$, which are separated by distances at most 2^j . By the hypothesis, such a triangle has a filling of area $O(2^{\alpha j})$. Similarly, the bigon can be filled at cost $O(n^\alpha)$. Finally, this is a template for w , so the cost to transform w to the boundary word is $O(n)$. Summing all the contributions, we find that $\delta_H(w) = O(n^\alpha)$. \square

Such a construction is used in [Gro93, 5.A''₃], [Gro96], and [LP04].

A third application is a proof that $SL(n; \mathbb{Z})$ has an exponential Dehn function. It is straightforward to show that the injectivity radius of $z \in \mathcal{M}$ shrinks exponentially quickly as $z \rightarrow \infty$; that is, that there is a c such that if $x, y \in \mathcal{E}$, $d_{\mathcal{E}}(I, x) \leq r$, and $d_{\mathcal{E}}(x, y) \leq e^{-cr}$, then $d_{\Gamma}(\rho(x), \rho(y)) \leq c$, where $\rho : \mathcal{E} \rightarrow \Gamma$ is the map from Sec. 2.4. If α is a curve of length $\ell = \ell_c(\alpha)$ in the thick part of \mathcal{E} , let $D^2(t) = [0, t] \times [0, t]$ and let $f : D^2(\ell) \rightarrow \mathcal{E}$ be a Lipschitz disc filling α ; this exists because \mathcal{E} is non-positively curved. We can in fact ensure that $\text{Lip}(f) \leq 2$, so that $d(I, f(x)) \leq 2\ell$ for all $x \in D^2(\ell)$. Let τ be a triangulation of $D^2(\ell)$ with $O(e^{4c\ell})$ triangles with side lengths at most $e^{-2c\ell}$. Label each vertex $v \in \tau^{(0)}$ by $\rho(f(v))$.

Each triangle in the template τ is labeled by three elements of Γ which are no more than c apart, so each ω -triangle can be filled with a disc of area $\delta_{\Gamma}(3c)$. This gives a filling of α with area $O(e^{2c\ell})$, as desired.

The proof is based on the fact that if two points in \mathcal{E} are exponentially close, then the corresponding elements of Γ are a bounded distance apart. The key to our proof that $SL(p; \mathbb{Z})$ has a quadratic Dehn function is that points in \mathcal{E} which are farther apart also satisfy certain relationships, so we can find a more efficient filling

by using a template with larger triangles. In a previous paper [You], we proved a quartic filling inequality for $SL(n; \mathbb{Z})$ by using a template with $O(\ell^2)$ triangles of side length at most 1. In this paper, we will improve this to a quadratic filling inequality by using a template with larger cells.

3.1. Adaptive triangulations. We claim that if w is a word in Σ which represents the identity, then there is a template τ for w such that all large triangles in τ have labels contained in a translate of $U(j, p - j)$ for some j . Furthermore, τ can be constructed with relatively few triangles. Specifically:

Proposition 3.2. *Let $p \in \mathbb{Z}$ and $p \geq 2$, and let $\Sigma, \Gamma, \mathcal{E}$, etc. be as in Sec. 2. There is a c such that if $w = w_1 \dots w_\ell$ is a word in Σ which represents the identity, then there is a template for w such that*

- (1) *If $g_1, g_2, g_3 \in \Gamma$ are the labels of a triangle in the template, then either $d_\Gamma(g_i, g_j) \leq 2c$ for all i, j or there is a k such that $g_i^{-1}g_j \in U(k, p - k)$ for all i, j .*
- (2) *τ has $O(\ell^2)$ triangles, and if the i th triangle of τ has vertices labeled (g_{i1}, g_{i2}, g_{i3}) , then*

$$\sum_i (d_\Gamma(g_{i1}, g_{i2}) + d_\Gamma(g_{i1}, g_{i3}) + d_\Gamma(g_{i2}, g_{i3}))^2 = O(\ell^2).$$

Similarly, if the i th edge of τ has vertices labeled h_{i1}, h_{i2} , then

$$\sum_i d_\Gamma(h_{i1}, h_{i2})^2 = O(\ell^2).$$

The basic technique is the same as the construction in the previous section; we start with a filling of w by a Lipschitz disc $f : D^2 \rightarrow \mathcal{E}$, then construct a template for w by triangulating the disc and labelling its vertices using ρ . We ensure that properties 1 and 2 hold by carefully controlling the lengths of edges. If edges are too long, then property 1 will not hold; on the other hand, if edges are too short, then there will be too many triangles, and 2 will not hold. Corollary 2.11 says that the length necessary for property 1 to hold varies based on where f lies in \mathcal{M} , so we will construct a triangulation with varying side lengths.

Proposition 3.3. *Let $t = 2^k$, $k \geq 0$, let $D^2(t) = [0, t] \times [0, t]$, and let $h : D^2(t) \rightarrow \mathbb{R}$ be a 1-Lipschitz function such that $h(x) \geq 1$ for all x . There is a triangulation τ_h of $D^2(t)$ such that*

- (1) *All vertices of τ_h are lattice points.*
- (2) *If x and y are connected by an edge of τ_h , then*

$$\frac{\min\{h(x)/4, t\}}{2} \leq d(x, y) \leq \sqrt{2}h(x).$$

- (3) *No more than 32 triangles meet at any vertex.*
- (4)

$$\sum_{\Delta \in \tau_h} (\text{perim } \Delta)^2 \leq 1152t^2$$

Furthermore, the number of triangles in τ_h is at most $32t^2$.

Proof. As in the construction of the Whitney decomposition, we will construct a decomposition of $D^2(t)$ into *dyadic squares*, that is, squares of the form

$$S_{i,j,s} := [i2^s, (i+1)2^s] \times [j2^s, (j+1)2^s]$$

for some $i, j, s \in \mathbb{Z}$, $s \geq 0$. Let \mathcal{D}_t be the set of dyadic squares contained in $D^2(t)$.

If S is a dyadic square, let $\sigma(S)$ be its side length and let $a(S)$ be the smallest dyadic square that strictly contains it, so that

$$a(S_{i,j,s}) = S_{\lfloor \frac{i}{2} \rfloor, \lfloor \frac{j}{2} \rfloor, s+1}.$$

If S and T are dyadic squares whose interiors intersect, then either $S \subset T$ and $T = a^k(S)$ for some k or vice versa.

Let

$$U_0 := \{S \mid S \in \mathcal{D}_t \text{ and } h(x) \geq \sigma(S) \text{ for all } x \in S\}$$

and

$$U := \{S \mid S \in U_0 \text{ and } a(S) \notin U_0\}.$$

We claim that U is a cover of $D^2(t)$ by squares which overlap only along their edges. If $x \in D^2(t)$, then $x \in S$ for some $S \in \mathcal{D}_t$ with $\sigma(S) = 1$. Since $h(z) \geq 1$ for all $z \in D^2(t)$, we know that $S \in U_0$; we claim that $a^n(S) \in U$ for some n . Indeed, if n is the largest integer such that $a^n(S) \in U_0$, then $a^n(S) \in U$. This c exists because when $2^n > t$, then $a^n(S) \notin U_0$. Thus U is a cover of $D^2(t)$.

Furthermore, squares in U intersect only along their edges. Let $S, T \in U$ be such that $\text{int } S \cap \text{int } T \neq \emptyset$ and $S \neq T$. In this case, we have $S \subsetneq T$ or $T \subsetneq S$; assume that $S \subsetneq T$. Then there is a n such that $a^n(S) = T$. By the definition of U , we know that $a(S) \notin U_0$, but this is impossible, since the fact that $T \in U_0$ implies that any dyadic square contained in T , including $a(S)$, is also in U_0 .

Two adjacent squares in U need not intersect along an entire edge, so U is generally not a polyhedron. To fix this, we subdivide the edges of each square so that two distinct polygons in U intersect either in a vertex, in an edge, or not at all; call the resulting polyhedron U' . By replacing each n -gon in U' with $n-2$ triangles, we obtain a triangulation, which we denote τ_h . We claim that this τ_h satisfies the required properties. The first property is clear; the vertices of any dyadic square are lattice points by definition.

For the second and third property, we will first show that if $x \in D^2(t)$ and $x \in S$ for some dyadic square $S \in U$, then

$$\frac{\min\{h(x)/4, t\}}{2} \leq \sigma(S) \leq h(x).$$

The inequality $\sigma(S) \leq h(x)$ follows from the definition of U . For the other inequality, let s be such that

$$2^s \leq \min\left\{\frac{h(x)}{4}, t\right\} < 2^{s+1},$$

and let S_0 be a dyadic square of side length 2^s such that $x \in S_0$. If $y \in S_0$, then $d(x, y) \leq 2^{s+1}$, so $h(y) \geq h(x) - 2^{s+1}$. Since $h(x) \geq 4 \cdot 2^s$, we have $h(y) \geq 2^{s+1} \geq \sigma(S_0)$, so $S_0 \in U_0$. Consequently, any square S in U must also contain S_0 , so

$$\sigma(S) \geq 2^s \geq \frac{\min\{h(x)/4, t\}}{2}$$

as desired. This implies property 2, because if e is an edge of τ_h , there is an $S \in U$ such that $e \subset S$ and $\sigma(S) \leq \ell_c(e) \leq \sqrt{2}\sigma(S)$.

Say that S and T are adjacent squares in U with $\sigma(S) \geq \sigma(T)$. By the above, we know that if $x \in S \cap T$, then $\sigma(S) \leq h(x)$ and

$$\frac{\min\{h(x)/4, t\}}{2} \leq \sigma(T),$$

so

$$\sigma(S) \leq 8\sigma(T).$$

This implies that a polygon in U' has at most 32 sides and that each vertex in τ_H has degree at most 128; this is property 3.

Finally, since U is a cover of $D^2(t)$ by squares with disjoint interiors,

$$\sum_{S \in U} \sigma(S)^2 = t^2.$$

Each square S in U corresponds to at most c triangles in τ_h , each of which has perimeter at most $6\sigma(S)$, so

$$\sum_{\Delta \in \tau_h} (\text{perim } \Delta)^2 \leq 36 \cdot 32t^2 = 1152t^2$$

as desired. Furthermore, U contains at most t^2 squares, so τ contains at most $32t^2$ triangles. \square

Proof of Prop. 3.2. Let $w(i) = w_1 \dots w_i$. Let $\alpha : [0, \ell] \rightarrow \mathcal{E}$ be the curve corresponding to w , parameterized so that $\alpha(i) = [w(i)]_{\mathcal{E}}$. If c_{Σ} is as in Sec. 2.3, then α is c_{Σ} -Lipschitz. Let $t = 2^k$ be the smallest power of 2 larger than ℓ , and let $\alpha' : [0, t] \rightarrow \mathcal{E}$

$$\alpha'(x) = \begin{cases} \alpha(x) & \text{if } x \leq \ell \\ [I]_{\mathcal{E}} & \text{otherwise.} \end{cases}$$

Since \mathcal{E} is non-positively curved, we can use geodesics to fill α' . If $x, y \in \mathcal{E}$, let $\gamma_{x,y} : [0, 1] \rightarrow \mathcal{E}$ be a geodesic parameterized so that $\gamma_{x,y}(0) = x$, $\gamma_{x,y}(1) = y$, and $\gamma_{x,y}$ has constant speed. We can define a homotopy $f : [0, t] \times [0, t] \rightarrow \mathcal{E}$ by

$$f(x, y) = \gamma_{\alpha'(x), \alpha'(0)}(y/t);$$

this sends three sides of $D^2(t)$ to $[I]_{\mathcal{E}}$ and is a filling of α . Since \mathcal{E} is non-positively curved, this map is $2c_{\Sigma}$ -Lipschitz and has area $O(\ell^2)$.

Let c' be the constant in Cor. 2.11. Let $r_0 : \mathcal{E} \rightarrow \mathbb{R}$ be

$$r_0(x) = \frac{d_{\mathcal{M}}([x]_{\mathcal{M}}, [I]_{\mathcal{M}})}{2p^2} - c',$$

and let $r : D^2(t) \rightarrow \mathbb{R}$ be

$$r(x) = \max\{1, \frac{r_0(x)}{4c_{\Sigma}}\},$$

This function is 1-Lipschitz. Let τ_r be the triangulation constructed in Prop. 3.3. Let τ be τ_r , with orientations on edges chosen arbitrarily. If v is an interior vertex of τ , label it $\rho(f(v))$. If $(i, 0)$ is a boundary vertex on the side of $D^2(t)$ corresponding to α' and $i \leq \ell$, label it by $w(i)$. Otherwise, label it I . If v is a vertex, let $g_v \in \Gamma$ be its labeling and note that $f(v) \in g_v \mathcal{S}$.

If x is a lattice point on the boundary of $D^2(t)$, then $f(x) = [I]_{\mathcal{M}}$ and so $r(x) = 1$. In particular, each lattice point on the boundary of $D^2(t)$ is a vertex of τ_0 , so the boundary of τ_0 is a $4t$ -gon with vertices labeled $I, w(1), \dots, w(n-1), I, \dots, I$. We identify vertices labeled I and remove self-edges to get a template τ for w .

If x_1 and x_2 are two adjacent vertices labeled g_1 and g_2 , then Prop. 3.3 implies that $d(x_1, x_2) \leq 2r(x_1)$ and

$$d_{\mathcal{E}}(f(x_1), f(x_2)) \leq 4c_{\Sigma}r(x_1) = \max\{4c_{\Sigma}, r_0(f(x_1))\}.$$

If

$$(2) \quad 4c_\Sigma \leq r_0(f(x_1)),$$

then

$$d_{\mathcal{E}}(f(x_1), f(x_2)) \leq r_0(f(x_1)),$$

so by Cor. 2.11, $g_1 x^{-1} g_2 \in U(j, p-j)$ for some j . Since j depends only on x_1 , if x_1, x_2 , and x_3 form a triangle, x_3 is labeled by g_3 , and (2) holds, then $g_1 x^{-1} g_3 \in U(j, p-j)$ as well, and thus $g_2 x^{-1} g_3 \in U(j, p-j)$.

Otherwise,

$$(3) \quad r_0(f(x_1)) \leq 4c_\Sigma,$$

so $d_{\mathcal{E}}(f(x_1), f(x_2)) \leq 4c_\Sigma$ and

$$d_{\mathcal{M}}([f(x_1)]_{\mathcal{M}}, [I]_{\mathcal{M}}) \leq 2p^2(c' + 4c_\Sigma).$$

In this case, $f(x_1)$ and $f(x_2)$ are both in the thick part of \mathcal{E} , and since

$$d_{\mathcal{E}}(f(x_1), f(x_2)) < 1,$$

there is a c'' independent of x_1 and x_2 such that $d_\Gamma(g_1, g_2) \leq c''$. As before, if x_1, x_2 , and x_3 form a triangle and (3) holds, then $d_\Gamma(g_1, g_3) \leq c''$, so $d_\Gamma(g_2, g_3) \leq 2c''$. This proves part 1 of 3.2.

To prove part 2, we will show that the distance between pairs of labels is essentially the same as the distance between the corresponding vertices in $D^2(t)$. This follows from Thm. 2.3. If (v_1, v_2) is an edge of τ and v_i is labeled by g_i , we know that

$$d_{\mathcal{M}}([f(v_i)]_{\mathcal{M}}, [I]_{\mathcal{M}}) = O(d_{D^2}(v_1, v_2)).$$

By Thm. 2.7,

$$d_{\mathcal{E}}([g_i]_{\mathcal{E}}, f(v_i)) = d_{\mathcal{M}}([I]_{\mathcal{M}}, [f(v_i)]_{\mathcal{M}}) + O(1) = O(d_{D^2}(v_1, v_2)),$$

so

$$\begin{aligned} d_{\mathcal{E}}([g_1]_{\mathcal{E}}, [g_2]_{\mathcal{E}}) &\leq d_{\mathcal{E}}([g_1]_{\mathcal{E}}, f(v_1)) + d_{\mathcal{E}}(f(v_1), f(v_2)) + d_{\mathcal{E}}(f(v_2), [g_2]_{\mathcal{E}}) \\ &= O(d_{D^2}(v_1, v_2)). \end{aligned}$$

By Thm. 2.3,

$$d_\Gamma(g_1, g_2) = O(d_{D^2}(v_1, v_2))$$

as well. Part 1 of Prop. 3.2 follows from this bound and the bounds in Prop. 3.3. This proves Prop. 3.2. \square

Note that it is not necessary that $p \geq 5$ for this template to exist. In fact, a suitable generalization of the proposition should hold for any lattice in a semisimple Lie group.

In the next section, we will fill this template. We will construct a normal form ω , and show that triangles and bigons whose edges are words in this normal form can be filled efficiently. Indeed, we will show that these triangles can be filled by discs with quadratically large area; by Lemma 2, this will give a quadratic filling of w .

4. THE COMBINATORIAL STEP: FILLING THE TEMPLATE

In the previous section, we constructed a template for a filling of a word w which represents the identity in $SL(p; \mathbb{Z})$. In this section, we will use this to prove Theorem 1.2. The first thing we need is a normal form ω for $SL(p; \mathbb{Z})$; we construct this in Section 4.1. The template from the previous section then allows us to reduce the problem of filling w with a disc of quadratic area to the problem of filling ω -triangles with vertices in a parabolic subgroup by discs of quadratic area. We construct these discs inductively, by reducing the problem of filling such an ω -triangle to the problem of finding relative fillings for words in $SL(q; \mathbb{Z})$, $q < p$ (that is, fillings of words in $SL(q; \mathbb{Z})$ by discs in $SL(p; \mathbb{Z})$).

Let $\Sigma_q := \Sigma \cap SL(q; \mathbb{Z})$ and $\Sigma_S := \Sigma \cap SL(S; \mathbb{Z})$; likewise, let $\widehat{\Sigma}_q := \widehat{\Sigma} \cap SL(q; \mathbb{Z})$ and $\widehat{\Sigma}_S := \widehat{\Sigma} \cap SL(S; \mathbb{Z})$.

In Sec. 4.1, we construct the normal form ω . In Sec. 4.2, we construct fillings of analogues of the Steinberg relations (see Sec. 2.3); these are our basic tools for manipulating words in $\widehat{\Sigma}$. In Sec. 4.3, we use these tools to reduce the problem of filling ω -triangles with vertices in a parabolic subgroup of $U(s_1, \dots, s_k) \subset SL(q; \mathbb{Z})$, $q \leq p$ to the problem of filling words in Σ_{s_i} and words in $\widehat{\Sigma}_2$, except for the case that the ω -triangle has vertices in $U(p-1, 1)$. In Sec. 4.4, we consider the case of $U(p-1, 1)$. Together, Secs. 3.1, 4.3, and 4.4 reduce the problem of filling words in Σ_q when $q \leq p$ to the problem of filling words in Σ_{q_i} where $3 \leq q_1, \dots, q_k < q$ and words in $\widehat{\Sigma}_2$. In Sec. 4.5, we find quadratic fillings of words in $\widehat{\Sigma}_2$. Finally, in Sec. 4.6, we bring all of these tools together to prove Thm. 1.2.

Throughout this section, p will be an integer which is at least 5.

4.1. Constructing a normal form. We first construct a normal form $\omega : \Gamma \rightarrow \Sigma^*$ for Γ . Let $g \in \Gamma$ and let $P = U(S_1, \dots, S_k) \in \mathcal{P}$ be the unique minimal $P \in \mathcal{P}$ containing g . Then g is a block-upper-triangular matrix which can be written as a product

$$(4) \quad g = \begin{pmatrix} m_1 & V_{12} & \dots & V_{1k} \\ 0 & m_2 & \dots & V_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & m_k \end{pmatrix} d,$$

where the i th block of the matrix corresponds to S_i . Here, $V_{i,j} \in \mathbb{Z}^{S_i} \otimes \mathbb{Z}^{S_j}$, and $d \in D$ is a diagonal matrix chosen so that $\det m_i = 1$. If $P = \Gamma$, then there is only one block, and we take $m_1 = g$, and $d = I$. If $V \in \mathbb{Z}^{S_i} \otimes \mathbb{Z}^{S_j}$, let $u_{ij}(V) = u_{S_i, S_j}(V)$. We can write g as a product:

$$\gamma_i := \left(\prod_{j=1}^{i-1} u_{ji}(V_{ji}) \right) m_i$$

$$g = \gamma_k \dots \gamma_1 d.$$

We will construct $\omega(g)$ by replacing the terms in this product decomposition with words. When $\#S_k \geq 3$, we can use Thm. 2.3 to replace the m_k by words in $\Sigma \cap SL(S_k; \mathbb{Z})$, but the theorem does not apply when $\#S_k = 2$ because $SL(2; \mathbb{Z})$ is exponentially distorted inside $SL(p; \mathbb{Z})$. We thus use a variant of the Lubotzky-Mozes-Ragunathan theorem to write m_k as a word in $\widehat{\Sigma} \cap SL(S_k; \mathbb{Z})$.

Proposition 4.1 ([LMR93]). *There is a constant c such that for all $g \in SL(k; \mathbb{Z})$, there is a word $w \in \widehat{\Sigma}^*$ which represents g and has length*

$$\widehat{\ell}(w) \leq c \log \|g\|_2.$$

For $i = 1, \dots, k$, let $\widehat{m}_k \in \widehat{\Sigma}^*$ be a word representing m_k as in Prop. 4.1. For $1 \leq i < j \leq k$ and $V \in \mathbb{Z}^{S_i} \otimes \mathbb{Z}^{S_j}$, let

$$\widehat{v}(V; S_i, S_j) := \prod_{a \in S_i, b \in S_j} e_{ab}(x_{ab}),$$

where x_{ab} is the (a, b) -coefficient of V .

Let

$$\begin{aligned} \widehat{\gamma}_i(g) &:= \left(\prod_{j=i+1}^k \widehat{v}(V_{ij}; S_i, S_j) \right) \widehat{m}_j \\ \omega_0(g) &= \widehat{\gamma}_k(g) \dots \widehat{\gamma}_1(g) d. \end{aligned}$$

This is a word in $\widehat{\Sigma}$ which represents g , and we let $\omega(g) = \lambda(\omega_0(g))$, where λ is the map defined in Sec. 2.3 which replaces each letter $e_{ab}(x)$ with the word $\widehat{e}_{ab}(x)$. It is straightforward to show that there is a constant c_ω independent of g such that $\ell_w(\omega(g)) \leq c_\omega d_\Gamma(I, g)$.

4.2. The shortened Steinberg relations. In this section, we will develop methods for filling simple words in $\widehat{\Sigma}$, based on the Steinberg relations.

The key to the methods in this section is the group

$$H_{S,T} = (\mathbb{R}^{s-1} \times \mathbb{R}^{t-1}) \ltimes (\mathbb{R}^S \otimes \mathbb{R}^T)$$

from Section 2.3, which we used to construct \widehat{e}_{ij} . This group has the key properties that it is contained in the thick part of G , and when either S or T is large enough, then $H_{S,T}$ has quadratic Dehn function. The quadratic Dehn function is a special case of a theorem of de Cornulier and Tessaera [dC08]:

Theorem 4.2. *If $s = \#S \geq 3$ or $t = \#T \geq 3$, then $H_{S,T}$ has a quadratic Dehn function.*

Proof. The groups $H_{S,T}$ and $H_{T,S}$ are isomorphic, so we may assume that $s \geq 3$. By a change of basis, we may assume that $\mathbb{R}^{s-1} \subset SL(S)$ and $\mathbb{R}^{t-1} \subset SL(T)$ are the subgroups of diagonal matrices with positive coefficients.

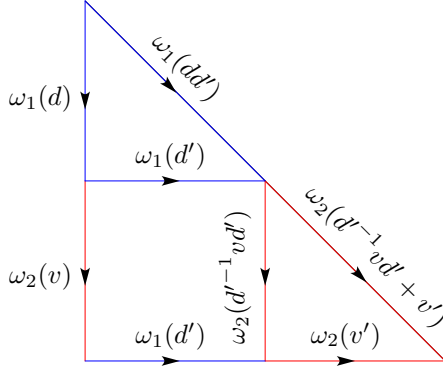
To prove that $H_{S,T}$ has a quadratic Dehn function, we use its semidirect product structure. Each element of $H_{S,T}$ can be written uniquely as a product dv of a $d \in \mathbb{R}^{s-1} \times \mathbb{R}^{t-1}$ and $v \in \mathbb{R}^S \otimes \mathbb{R}^T$, so if ω_1 and ω_2 are normal forms for $\mathbb{R}^{s-1} \times \mathbb{R}^{t-1}$ and $\mathbb{R}^S \otimes \mathbb{R}^T$, we get a normal form for H by letting $\omega_H(dv) = \omega_1(d)\omega_2(v)$. By Prop. 3.1, it suffices to show that curves of the form $z = \omega_H(dv)\omega_H(d'v')\omega_H(dvd'v')^{-1}$ have quadratic fillings. These can be rewritten in the form

$$\omega_1(d)\omega_2(v)\omega_1(d')\omega_2(v')\omega_2(d'^{-1}vd' + v')^{-1}\omega_1(dd')^{-1},$$

where multiplication in $\mathbb{R}^S \otimes \mathbb{R}^T$ is written additively. This curve forms the boundary of the triangle in Fig. 3, so we can fill it by filling the curves

$$(5) \quad \omega_2(v)\omega_1(d')\omega_2(d'^{-1}vd')^{-1}\omega_1(d')^{-1},$$

$$(6) \quad \omega_1(d)\omega_1(d')\omega_1(dd')^{-1},$$

FIGURE 3. A quadratic filling of z .

and

$$(7) \quad \omega_2(d'^{-1}vd')\omega_2(v')\omega_2((d'^{-1}vd') + v')^{-1}$$

which form the boundaries of the cells of Fig. 3.

First, we define ω_1 and ω_2 . We let $\omega_1(d)$ be the map $t \mapsto d^t$ for $t \in [0, 1]$; by abuse of notation, we also call this curve d . We will construct $\omega_2(v)$ the same way that we constructed $\hat{u}(v; S, T)$. Let $D_i(\lambda) \in SL(S)$ be the diagonal matrix such that $D_i(\lambda)z_i = \lambda z_i$ and $D_i(\lambda)z_j = \lambda^{-\frac{1}{s-1}}z_j$. If $v \in \mathbb{R}^S \otimes \mathbb{R}^T$, let v_{ij} be the (i, j) -component of v , and if $x \in \mathbb{R}$, let $\bar{x} = \max\{1, |x|\}$. We define

$$\omega_2(v) := \prod_{i,j} \gamma_{ij}(v_{ij}),$$

where

$$\gamma_{ij}(x) = D_i(\bar{x})e_{ij}(x/\bar{x})D_i(\bar{x})^{-1}$$

is a curve representing $e_{ij}(x)$.

Now we fill (5). It suffices to consider the case that $v = xz_i \otimes z_j$; the general case is a combination of such cases. Let λ be such that $d'^{-1}vd' = \lambda v$. We need to fill

$$w_0 = \gamma_{ij}(x)d'\gamma_{ij}(\lambda x)^{-1}d'^{-1}.$$

We can conjugate w_0 by $D_i(\bar{x})$ and collect the diagonal matrices to get $w_1 = e_{ij}(x_1)De_{ij}(x_2)D^{-1}$, where $x_1 = x/\bar{x}$, $x_2 = \lambda x/\bar{\lambda}x$, and $D = D_i(\bar{x})^{-1}d'D_i(\bar{\lambda}x)$. Conjugating by $D_i(\bar{x})$ is just a change of basepoint, so it has no cost, and because $\mathbb{R}^{s-1} \times \mathbb{R}^{t-1}$ has a quadratic Dehn function, collecting diagonal matrices has cost $O(\ell_c(w_0)^2)$.

This has a thin rectangle as a filling: the map $\beta : [0, \ell_c(D)] \times [0, 1] \rightarrow H_{S,T}$

$$\beta(a, b) = e_{ij}(bx_1)D^a$$

is a distance-decreasing map with boundary w_1 . This gives a filling of (5) with quadratic area, as desired.

The curve (6) can be filled quadratically because $\mathbb{R}^{s-1} \times \mathbb{R}^{t-1}$ has a quadratic Dehn function. The curve (7) is more complicated. Because $\omega_2(v)$ is composed of

curves $\gamma_{ij}(x)$, it suffices to find fillings of combinations of the $\gamma_{ij}(x)$. Namely, if $\overline{\log x} = \max\{1, \log x\}$,

$$(8) \quad \delta(\gamma_{ij}(x)\gamma_{ij}(y), \gamma_{ij}(x+y)) = O((\overline{\log x} + \overline{\log y})^2)$$

for all x, y and

$$(9) \quad \delta([\gamma_{ij}(x), \gamma_{kl}(y)]) = O((\overline{\log x} + \overline{\log y})^2)$$

when $i \neq k$ or $j \neq l$.

Recall that the group $\text{Sol}_{2s-1} = \mathbb{R}^{s-1} \ltimes \mathbb{R}^s$ has a quadratic Dehn function when $s \geq 3$ [Gro93, 5.A.9]. $H_{S,T}$ contains several copies of Sol_{2s-1} ; if $j_1, \dots, j_t \in \{1, \dots, t\}$, then the subgroup of $H_{S,T}$ generated over \mathbb{R} by $D_1, \dots, D_s, e_{1j_1}, \dots, e_{sj_s}$ is isomorphic to Sol_{2s-1} . In particular, curves of the form (8) are curves in a copy of Sol_{2s-1} and when $i \neq k$, curves of the form (9) are curves in a copy of Sol_{2s-1} . Such curves have quadratic fillings.

It remains to consider the case that $i = k$ and show that

$$\delta([\gamma_{ij}(x), \gamma_{ik}(y)]) = O((\overline{\log |x|} + \overline{\log |y|})^2).$$

Assume without loss of generality that $x \leq y$; if $y \leq 1$, then the curve has bounded length, so we may take $y > 1$. Since $x \leq y$, we have $\lambda_{ij}(x/y) = e_{ij}(x/y)$, and the curve

$$\delta(\gamma_{ij}(x), D_i(y)e_{ij}(x/y)D_i(y)^{-1})$$

is a curve of the form (5). The techniques used to fill (5) show that it has a filling of area $O((\overline{\log |x|} + \overline{\log |y|})^2)$. thus since

$$\gamma_{ik}(y) = D_i(y)e_{ik}(1)D_i(y)^{-1},$$

we have

$$\delta(\gamma_{ij}(x)\gamma_{ik}(y), D_i(y)u(x/y z_i \otimes z_j + z_i \otimes z_k)D_i(y)^{-1}) = O((\overline{\log |x|} + \overline{\log |y|})^2)$$

as well. Similarly,

$$\delta(\gamma_{ij}(y)\gamma_{ik}(x), D_i(y)u(z_i \otimes z_k + x/y z_i \otimes z_j)D_i(y)^{-1}) = O((\overline{\log |x|} + \overline{\log |y|})^2),$$

so

$$\delta([\gamma_{ij}(x), \gamma_{ik}(y)]) = O((\overline{\log |x|} + \overline{\log |y|})^2),$$

as desired. \square

Recall that we defined the words $\hat{e}_{ij}(x)$ as approximations of curves $\hat{u}(v; S, T)$ in the $H_{S,T}$. We can use the fact that $H_{S,T}$ has quadratic Dehn function to manipulate these curves. In the next lemma, we find fillings for words representing conjugations of $\hat{u}(V; S, T)$. Let $\Sigma_S := \Sigma \cap SL(S; \mathbb{Z})$ and $\Sigma_T := \Sigma \cap SL(T; \mathbb{Z})$. These are generating sets for $SL(S; \mathbb{Z})$ and $SL(T; \mathbb{Z})$.

Lemma 4.3. *Let $0 < \epsilon < 1/2$ be sufficiently small that $H_{S,T} \subset G(\epsilon)$. If $\#S \geq 3$ and $\#T \geq 2$ or vice versa, γ is a word in $(\Sigma_S \cup \Sigma_T)^*$ representing $(M, N) \in SL(S; \mathbb{Z}) \times SL(T; \mathbb{Z})$, and $V \in \mathbb{R}^S \otimes \mathbb{R}^T$, then*

$$\delta_{\mathcal{E}(\epsilon)}(\gamma \hat{u}(V; S, T) \gamma^{-1}, \hat{u}(MVN^{-1}; S, T)) = O((\ell_w(\gamma) + \log(\|V\|_2 + 2))^2).$$

Proof. In this proof, we will write $u(V; S, T)$ and $\hat{u}(V; S, T)$ as $u(V)$ and $\hat{u}(V)$, leaving S and T implicit. Let

$$\omega := \gamma \hat{u}(V) \gamma^{-1} \hat{u}(MVN^{-1})^{-1};$$

this is a closed curve in $G(\epsilon)$ of length $O((\ell_w(\gamma) + \log(\|V\|_2 + 2))^2)$.

We first consider the case that $V = xv_i \otimes w_j$ and $\gamma \in \Sigma_T^*$. In this case, $M = I$; and $\gamma\hat{u}(V)\gamma^{-1}$ and $\hat{u}(VN^{-1})$ are both words in the alphabet

$$\Sigma_F := \{A_i^x \mid x \in \mathbb{R}\} \cup \{u(W) \mid W \in \mathbb{R}^S \otimes \mathbb{R}^T\} \cup \Sigma_T.$$

Furthermore, Σ_F generates the group

$$\begin{aligned} F &:= \left\{ \begin{pmatrix} \prod_i A_i^{x_i} & W \\ 0 & D \end{pmatrix} \middle| x_i \in \mathbb{R}, D \in SL(T; \mathbb{Z}), W \in \mathbb{R}^S \otimes \mathbb{R}^T \right\} \\ &= (\mathbb{R}^{s-1} \times SL(T; \mathbb{Z})) \ltimes (\mathbb{R}^S \otimes \mathbb{R}^T), \end{aligned}$$

and $F \subset G(\epsilon)$, so words in Σ_F^* correspond to curves in $\mathcal{E}(\epsilon)$.

Words in Σ_F^* satisfy certain relations which correspond to discs in $\mathcal{E}(\epsilon)$. In particular, note that if $\sigma \in \Sigma_T$, $|x| \leq 1$, and $\|W\|_2 \leq 1$, then

$$(10) \quad [\sigma, A_k^x]$$

and

$$(11) \quad \sigma u(W) \sigma^{-1} u(W \sigma^{-1})^{-1}$$

are both closed curves of bounded length and thus have bounded filling areas. We can think of them as “relations” in F .

Let $\lambda_i > 1$ be such that $\lambda_i v_i = A_i v_i$ for $i = 1, \dots, s$. Let $C = \log_{\min_k \{\lambda_k\}}(p+1)$, and let $z = C\ell_w(\gamma) + \max\{1, \log_{\lambda_i} |x|\}$. This choice of z ensures that

$$\|VN\|_2 \leq \lambda_i^z.$$

Indeed, it ensures that if $d_{SL(T; \mathbb{Z})}(I, N') \leq \ell_w(\gamma)$, then

$$\|VN'\|_2 \leq \lambda_i^z.$$

Furthermore, $z = O(\ell_c(\omega))$.

We will construct a homotopy which lies in $\mathcal{E}(\epsilon)$ and goes through the stages

$$\begin{aligned} \omega_1 &= \gamma\hat{u}(V)\gamma^{-1} \\ \omega_2 &= \gamma A_i^z u(\lambda_i^{-z} V) A_i^{-z} \gamma^{-1} \\ \omega_3 &= A_i^z \gamma u(\lambda_i^{-z} V) \gamma^{-1} A_i^{-z} \\ \omega_4 &= A_i^z u(\lambda_i^{-z} V N^{-1}) A_i^{-z} \\ \omega_5 &= \hat{u}(VN^{-1}). \end{aligned}$$

Each stage is a word in Σ_F^* and so corresponds to a curve in $\mathcal{E}(\epsilon)$.

We can construct a homotopy between ω_1 and ω_2 and between ω_4 and ω_5 using Thm. 4.2. We need to construct homotopies between ω_2 and ω_3 and between ω_3 and ω_4 .

We can transform ω_2 to ω_3 by applying (10) at most $O(\ell_c(\omega)^2)$ times. This corresponds to a homotopy with area $O(\ell_c(\omega)^2)$. Similarly, we can transform ω_3 to ω_4 by applying (11) at most $\ell_w(\gamma)$ times, corresponding to a homotopy of area $O(\ell_w(\gamma))$. Combining all of these homotopies, we find that

$$\delta_{\mathcal{E}(\epsilon)}(\gamma\hat{u}(V)\gamma^{-1}, \hat{u}(VN^{-1})) \leq O(\ell_c(\omega)^2).$$

as desired.

We can generalize to the case $V = \sum_{i,j} x_{ij} v_i \otimes w_j$ and $\gamma \in \Sigma_T^*$. By applying the case to each term of $\hat{u}(V)$, we obtain a homotopy of area $O(\ell_c(\omega)^2)$ from $\gamma \hat{u}(V) \gamma^{-1}$ to

$$\prod_{i,j} \hat{u}(x_{ij} v_i \otimes w_j N^{-1}).$$

This is a curve in $H_{S,T}$ of length $O(\ell_c(\omega))$ which connects I and $u(VN^{-1})$. By Thm. 4.2, there is a homotopy between this curve and $\hat{u}(VN^{-1})$ of area $O(\ell_c(\omega)^2)$.

When $\gamma \in \Sigma_S^*$, we instead let F be the group

$$F := \left\{ \begin{pmatrix} D & W \\ 0 & \prod_i B_i^{x_i} \end{pmatrix} \middle| x_i \in \mathbb{R}, D \in SL(S; \mathbb{Z}), W \in \mathbb{R}^S \otimes \mathbb{R}^T \right\} \\ = (SL(S; \mathbb{Z}) \times \mathbb{R}^{t-1}) \ltimes (\mathbb{R}^S \otimes \mathbb{R}^T).$$

Here, $\hat{u}(V)$ is not a word in F , but since $\#T \geq 2$, we can replace the A_i with the B_i in the construction of $\hat{u}(V)$. This results in shortcuts $\hat{u}'(V)$ in the alphabet

$$\{B_i^x \mid x \in \mathbb{R}\} \cup \{u(V) \mid V \in \mathbb{R}^S \otimes \mathbb{R}^T\}.$$

These are curves in $H_{S,T}$ which represent $u(V)$ and have length $O(\log \|V\|_2)$, so by Thm. 4.2, there is a homotopy of area $O((\log \|V\|_2)^2)$ between $\hat{u}'(V)$ and $\hat{u}(V)$.

Modifying the argument appropriately, we can show that when $\gamma \in \Sigma_S^*$,

$$\delta_{\mathcal{E}(\epsilon)}(\gamma \hat{u}'(V) \gamma^{-1}, \hat{u}'(MV)) = O(\ell_c(\omega)^2).$$

Replacing $\hat{u}'(V)$ with $\hat{u}(V)$ and $\hat{u}'(MV)$ with $\hat{u}(MV)$ adds area $O(\ell_c(\omega)^2)$, so

$$\delta_{\mathcal{E}(\epsilon)}(\gamma \hat{u}(V) \gamma^{-1}, \hat{u}(MV)) = O(\ell_c(\omega)^2).$$

If $\gamma \in (\Sigma_S \cup \Sigma_T)^*$, and $\gamma_S \in \Sigma_S^*$ and $\gamma_T \in \Sigma_T^*$ are the words obtained by deleting all the letters in Σ_T and Σ_S respectively, then $\delta_\Gamma(\gamma, \gamma_S \gamma_T) = O(\ell_c(\omega)^2)$. We can construct a homotopy from $\gamma \hat{u}(V) \gamma^{-1}$ to $\hat{u}(MVN^{-1})$ which goes through the steps

$$\begin{aligned} \gamma \hat{u}(V) \gamma^{-1} &\rightarrow \gamma_S \gamma_T \hat{u}(V) \gamma_T^{-1} \gamma_S^{-1} \\ &\rightarrow \gamma_S \hat{u}(VN^{-1}) \gamma_S^{-1} \\ &\rightarrow \hat{u}(MVN^{-1}). \end{aligned}$$

This homotopy has area $O(\ell_c(\omega)^2)$. \square

When we constructed $\hat{e}_{ij}(x)$, we made a choice of d for each pair (i, j) . The next lemma shows that this choice doesn't matter very much; different choices of d lead to curves which can be connected by a quadratic homotopy.

Lemma 4.4. *If $i \in S, S'$ and $j \in T, T'$, where $2 \leq \#S, \#S' \leq p-2$, then*

$$\delta_\Gamma(\hat{e}_{ij;S,T}(x), \hat{e}_{ij;S',T'}(x)) = O((\log |x|)^2).$$

In particular,

$$\delta_\Gamma(\hat{e}_{ij}(x), \hat{e}_{ij;S,T}(x)) = O((\log |x|)^2).$$

Proof. Let $V = xz_i \otimes z_j$. We first consider two special cases: the case that $S = S'$, and the case that $S \subset S'$, $\#S' \geq 3$, $T \subset T'$, and $\#T' \geq 2$.

Case 1: $S = S'$. In this case, $\hat{u}(V; S, T)$ and $\hat{u}(V; S', T')$ are both curves in H_{S, S^c} for S^c the complement of S . Since $p \geq 5$, this has quadratic Dehn function, so the lemma follows. In particular,

$$\delta_\Gamma(\hat{e}_{ij;S,T}(x), \hat{e}_{ij;S, \{j\}}(x)) = O((\log |x|)^2).$$

Case 2: $S \subset S'$, $\#S' \geq 3$, $T \subset T'$, and $\#T' \geq 2$. Let $\{A_i\}$ be as in the definition of $H_{S,T}$, with eigenvectors v_i and let $\{A'_i\} \in SL(S'; \mathbb{Z})$ be the set of independent commuting matrices used in defining $H_{S',T'}$. Recall that $\hat{u}(V; S, T)$ is the concatenation of curves γ_i of the form

$$\gamma_i = A_i^{c_i} u(x_i v_i \otimes z_j) A_i^{-c_i}$$

where $c_i \in \mathbb{R}$ and $|x_i| \leq 1$. Let α_i be a word in Σ representing A_i . There is a homotopy between γ_i and

$$\gamma'_i = \alpha_i^{\lfloor c_i \rfloor} u(\lambda_i^{c_i - \lfloor c_i \rfloor} x_i v_i \otimes z_j) \alpha_i^{-\lfloor c_i \rfloor}$$

of area $O(\log |x|)$. Since α_i is a word in $\Sigma_{S'}$, Lemma 4.3 implies that there is a homotopy of area $O((\log |x|)^2)$ between γ'_i and

$$\gamma''_i = \hat{u}(\lambda_i^{c_i} x_i v_i \otimes z_j; S', T').$$

Each of the γ''_i lie in $H_{S',T'}$, and the product of the elements they represent is $e_{ij}(x)$. Since $\hat{u}(V; S', T')$ also lies in $H_{S',T'}$ and $H_{S',T'}$ has quadratic Dehn function, this implies that

$$\delta_\Gamma(\hat{e}_{S,T}(V), \hat{e}_{S',T'}(V)) = O((\log |x|)^2),$$

as desired.

Combining these two cases proves the lemma. First, we construct a homotopy between $\hat{e}_{S,T}(V)$ and a word of the form $\hat{e}_{\{i,d\},\{j\}}(V)$. Let $d \in S$ be such that $d \neq i$. We can construct a homotopy going through the stages

$$\hat{e}_{S,T}(V) \rightarrow \hat{e}_{S,S^c}(V) \rightarrow \hat{e}_{\{i,d\},\{j\}}(V).$$

The first step uses case 1; the second step uses case 2, since $j \in S^c$.

We can use the same procedure to construct a homotopy between $\hat{e}_{S',T'}(V)$ and a word of the form $\hat{e}_{\{i,d'\},\{j\}}(V)$. If $d = d'$, we're done. Otherwise, we can use case 2 to construct homotopies between each word and $\hat{e}_{\{i,d,d'\},\{i,d,d'\}^c}(V)$. \square

One important use of this lemma is that if $\#S \geq 3$ and $i, j, d \in S$ are distinct, then the lemma lets us replace $\hat{e}_{ij}(x)$ by $\hat{e}_{ij;\{i,d\},\{j\}}(x)$ for a cost of $O((\log |x|)^2)$. This is a word in Σ_S . More generally,

Corollary 4.5. *If $i \in S, S'$ and $j \in T, T'$, where $2 \leq \#S, \#S' \leq p-2$, and $V \in \mathbb{R}^{S \cap S'} \otimes \mathbb{R}^{T \cap T'}$, then*

$$\delta_{\mathcal{E}(1/2)}(\hat{u}(V; S, T), \hat{u}(V; S', T')) = O((\log \|V\|)^2).$$

Proof. Note that $\hat{u}(V; S, T) = \hat{u}(V; S, S^c)$, so we may assume that $T = S^c$ and $T' = S'^c$, so $H_{S,T}$ and $H_{S',T'}$ have quadratic Dehn functions.

Let $V = \sum_{i,j} x_{ij} z_i \otimes z_j$. Note that $\prod_{i,j} \hat{u}(x_{ij} z_i \otimes z_j; S, T)$ is a curve in $H_{S,T}$ which represents $u(V; S, T)$. There is a homotopy in \mathcal{E} going through the stages:

$$\omega_1 = \hat{u}(V; S, T)$$

$$\omega_2 = \prod_{i,j} \hat{u}(x_{ij} z_i \otimes z_j; S, T)$$

$$\omega_3 = \prod_{i,j} \hat{u}(x_{ij} z_i \otimes z_j; S', T')$$

$$\omega_4 = \hat{u}(V; S', T').$$

Here, ω_1 and ω_2 are curves in $H_{S,T}$, so there is a quadratic-area homotopy from one to the other. Likewise, ω_3 and ω_4 are both curves in $H_{S',T'}$. We can use Lemma 4.4 to construct the homotopy from ω_2 to ω_3 .

These homotopies lie in the thick part of \mathcal{E} and have total area $O((\log |V|)^2)$. \square

Using these lemmas, we can give fillings for a wide variety of curves, including shortened versions of the Steinberg relations.

Lemma 4.6. *If $x, y \in \mathbb{Z} - \{0\}$, then*

- (1) *If $1 \leq i, j \leq p$ and $i \neq j$, then*

$$\delta_\Gamma(\widehat{e}_{ij}(x)\widehat{e}_{ij}(y), \widehat{e}_{ij}(x+y)) = O((\log |x| + \log |y|)^2).$$

In particular,

$$\delta_\Gamma(\widehat{e}_{ij}(x)\widehat{e}_{ij}(-x)) = \delta_\Gamma(\widehat{e}_{ij}(x)^{-1}, \widehat{e}_{ij}(-x)) = O((\log |x|)^2).$$

- (2) *If $1 \leq i, j, k \leq p$ and $i \neq j \neq k$, then*

$$\delta_\Gamma([\widehat{e}_{ij}(x), \widehat{e}_{jk}(y)], \widehat{e}_{ik}(xy)) = O((\log |x| + \log |y|)^2).$$

- (3) *If $1 \leq i, j, k, l \leq p$, $i \neq l$, and $j \neq k$*

$$\delta_\Gamma([\widehat{e}_{ij}(x), \widehat{e}_{kl}(y)]) = O((\log |x| + \log |y|)^2).$$

- (4) *Let $1 \leq i, j, k, l \leq p$, $i \neq j$, and $k \neq l$, and*

$$s_{ij} = e_{ji}^{-1} e_{ij} e_{ji}^{-1},$$

so that s_{ij} represents

$$\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \in SL(\{i, j\}; \mathbb{Z}).$$

Then

$$\delta_\Gamma(s_{ij}\widehat{e}_{kl}(x)s_{ij}^{-1}, \widehat{e}_{\sigma(k)\sigma(l)}(\tau(k, l)x)) = O((\log |x| + \log |y|)^2),$$

where σ is the permutation switching i and j , and $\tau(k, l) = -1$ if $k = i$ or $l = i$ and 1 otherwise.

- (5) *If $b = \text{diag}(b_1, \dots, b_p)$, then*

$$\delta_\Gamma(b\widehat{e}_{ij}(x)b^{-1}, \widehat{e}_{ij}(b_i b_j x)) = O(\log |x|^2).$$

Proof. For part 1, let $S = \{i, d\}$ and $T = S^c$. We can use Lemma 4.4 to replace

$$\widehat{e}_{ij}(x)\widehat{e}_{ij}(y)\widehat{e}_{ij}(x+y)^{-1}$$

by

$$\widehat{e}_{ij;S,T}(x)\widehat{e}_{ij;S,T}(y)\widehat{e}_{ij;S,T}(x+y)^{-1}$$

This is an approximation of a closed curve in H_{S,S^c} of length $O(\log |x| + \log |y|)$, which can be filled using Thm. 4.2.

For part 2, let $d \notin \{i, j, k\}$ and let $S = \{i, j, d\}$, so that $\widehat{e}_{ij;\{i,d\},\{j\}}(x)$ is a word in $SL(S; \mathbb{Z})$. We construct a homotopy going through the stages

$$\omega_0 = [\widehat{e}_{ij}(x), \widehat{e}_{jk}(y)]\widehat{e}_{ik}(xy)^{-1}$$

$$\omega_1 = [\widehat{e}_{ij;\{i,d\},\{j\}}(x), \widehat{u}(yz_j \otimes z_k; S, S^c)]\widehat{e}_{ik;S,S^c}(xy)^{-1}$$

$$\omega_2 = \widehat{u}((xyz_i + yz_j) \otimes z_k; S, S^c)\widehat{u}(yz_j \otimes z_k; S, S^c)^{-1}\widehat{u}(xyz_i \otimes z_k; S, S^c)^{-1}.$$

Here, we use Lemma 4.4 to construct a homotopy between ω_0 and ω_1 . The homotopy between ω_1 and ω_2 is an application of Lemma 4.3 with $\gamma = \widehat{e}_{ij;\{i,d\},\{j\}}(x)$ and

$V = yz_j \otimes z_k$. Finally, ω_2 is a curve in H_{S, S^c} with length $O(\log |x| + \log |y|)$, and thus has filling area $O((\log |x| + \log |y|)^2)$. The total area used is $O((\log |x| + \log |y|)^2)$.

For part 3, we let $S = \{i, k, d\}$, $T = \{j, l\}$, and use the same techniques to construct a homotopy going through the stages

$$\begin{aligned} & [\widehat{e}_{ij}(x), \widehat{e}_{kl}(y)] \\ & [\widehat{e}_{ij;S,T}(x), \widehat{e}_{kl;S,T}(y)] && \text{by Lem. 4.4} \\ & \varepsilon && \text{by Thm. 4.2,} \end{aligned}$$

where ε represents the empty word. This homotopy has area $O((\log |x| + \log |y|)^2)$.

Part 4 breaks into several cases depending on k and l . When i, j, k , and l are distinct, the result follows from part 3, since $s_{ij} = e_{ji}^{-1} e_{ij} e_{ji}^{-1}$, and we can use part 3 to commute each letter past $\widehat{e}_{kl}(x)$. If $k = i$ and $l \neq j$, let $d, d' \notin \{i, j, l\}$, $d \neq d'$, and let $S = \{i, j, d\}$ and $T = \{l, d'\}$. There is a homotopy from

$$s_{ij} \widehat{e}_{il}(x) s_{ij}^{-1} \widehat{e}_{jl}(-x)^{-1}$$

to

$$s_{ij} \widehat{u}(xz_i \otimes z_l; S, T) s_{ij}^{-1} \widehat{e}_{jl}(xz_j \otimes z_l)$$

of area $O((\log |x|)^2)$, and since $s_{ij} \in \Sigma_S^*$, the proposition follows by an application of Lemma 4.3. A similar argument applies to the cases $k = j$ and $l \neq i$; $k \neq i$ and $l = j$; and $k \neq j$ and $l = i$.

If $i = k$ and $j = l$, let $d \notin \{i, j\}$. There is a homotopy going through the stages

$$\begin{aligned} & s_{ij} \widehat{e}_{ij}(x) s_{ij}^{-1} \\ & s_{ij} [e_{id}, \widehat{e}_{dj}(x)] s_{ij}^{-1} && \text{by part (2)} \\ & [s_{ij} e_{id} s_{ij}^{-1}, s_{ij} \widehat{e}_{dj}(x) s_{ij}^{-1}] && \text{by free insertion} \\ & [e_{jd}^{-1}, \widehat{e}_{di}(x)] && \text{by previous cases} \\ & \widehat{e}_{ji}(-x) && \text{by part (2)} \end{aligned}$$

and this homotopy has area $O((\log |x|)^2)$. One can treat the case that $i = l$ and j_k the same way.

Since any diagonal matrix in Γ is the product of at most p elements s_{ij} , part 5 follows from part 4. \square

4.3. Reducing to smaller groups. In this section, we will apply the constructions in Lemma 4.6 to reduce the problem of filling an ω -triangle with vertices in $P = U(S_1, \dots, S_n)$ to the problem of filling loops in $SL(S_i; \mathbb{Z})$. The main caveat is that the methods in this section fails when $P = U(p-1, 1)$; this case will be left to Section 4.4.

Let $P = U(S_1, \dots, S_n; \mathbb{Z})$. Because the ω -triangles were constructed as products of the $\omega_0(g)$, which are words in $\widehat{\Sigma}$, we will work primarily with words in $\widehat{\Sigma}$. If w is a word in $\widehat{\Sigma}$, define $\widehat{\delta}(w) := \delta_\Gamma(\lambda(w))$, where λ is the map defined in Sec. 2.3 which replaces each letter $e_{ab}(x)$ with the word $\widehat{e}_{ab}(x)$.

Let $P^+ \subset P$ be the finite-index subgroup consisting of matrices in P whose diagonal blocks all have determinant 1, and let $\widehat{\Sigma}_{P^+} := \widehat{\Sigma} \cap P^+$. If w is a word in $\widehat{\Sigma} \cap P$ which represents the identity, we can modify it to get a word in $\widehat{\Sigma}_{P^+}$. First, we can use Lemma 4.6 to gather together all of the diagonal matrices at cost $O(\widehat{\ell}(w)^2)$; this results in a word dw' where d is the product of all the diagonal matrices in w and w' contains no diagonal matrices. Since w' represents an element of P^+ ,

$d' \in P^+$ as well. We decompose d' as a product $d_1 \dots d_n$ such that $d_i \in SL(S_i; \mathbb{Z})$. Let $f(w) = d_1 \dots d_n w'$; this is a word in $\widehat{\Sigma}_{P^+}$, and $\widehat{\ell}(f(w)) \leq \widehat{\ell}(w) + p$. Let $p_i : P^+ \rightarrow SL(S_i; \mathbb{Z})$ be the map projecting a block-upper triangular matrix to one of its diagonal blocks; this map extends to a map on $\widehat{\Sigma}_{P^+}^*$ which we also denote p_i . This map sends each letter of $f(w)$ either to itself or to the identity, so $p_i(w)$ is the word in $\widehat{\Sigma}_{S_i} := \widehat{\Sigma} \cap SL(S_i; \mathbb{Z})$ obtained by deleting all the letters of $f(w)$ except those in $\widehat{\Sigma}_{S_i}$; since w represented the identity, $p_i(w)$ also represents the identity.

The main goal of this section is to prove:

Proposition 4.7. *Let $P = U(S_1, \dots, S_k)$, where $\#S_i \leq p-2$ for all i . If $g_1, g_2, g_3 \in P$, where $g_1 g_2 g_3 = 1$, and*

$$w = \omega_0(g_1^{e_1})^{e_1} \omega_0(g_2^{e_2})^{e_2} \omega_0(g_3^{e_3})^{e_3},$$

where $e_i = \pm 1$, then

$$\widehat{\delta}(w, p_1(w) \dots p_k(w)) = O(\widehat{\ell}(w)^2).$$

In particular, this lemma implies that

$$\widehat{\delta}(w) \leq \sum_i \widehat{\delta}(p_i(w)) + O(\widehat{\ell}(w)^2).$$

Recall that if $g \in \Gamma$ and $Q = U(T_1, \dots, T_r)$ is the minimal element of \mathcal{P} containing g , then

$$\omega_0(g) = \widehat{\gamma}_r(g) \dots \widehat{\gamma}_1(g)d,$$

where $d \in D$,

$$\widehat{v}(V; T_i, T_j) := \prod_{a \in T_i, b \in T_j} e_{ab}(x_{ab}),$$

$$\widehat{\gamma}_i(g) := \left(\prod_{j=i+1}^r \widehat{v}(V_{ij}; T_i, T_j) \right) \widehat{m}_j,$$

and \widehat{m}_j is a word in $\widehat{\Sigma}_{T_j}^*$. If $g \in P$, then $Q \subset P$, and for each $1 \leq i \leq r$, there is an i' such that $T_i \subset S_{i'}$. As a consequence, \widehat{m}_i is a word in $\widehat{\Sigma}_{S_{i'}}$, and $\widehat{v}(V_{ij}; T_i, T_j)$ is either a word in $\widehat{\Sigma}_{S_{i'}}$ or can be written as $\widehat{v}(V_{ij}; S_{i'}, S_{j'})$. We can thus write $f(w)$ as a product of at most $3p^2 + 9p$ words so that each subword is either a word in $\widehat{\Sigma}_{S_i}$ for some i or a word of the form $\widehat{v}(V; S_i, S_j)$ for some i, j and $V \in \mathbb{R}^{S_i} \otimes \mathbb{R}^{S_j}$; henceforth, we will write $\widehat{v}_{ij}(V) := \widehat{v}(V; S_i, S_j)$. We will prove Prop. 4.7 by giving methods for collecting the subwords of the first type. This involves three main techniques: conjugating $\widehat{v}_{ij}(V)$ by a word in $\widehat{\Sigma}_{S_i}$ (Lemma 4.9), commuting words in $\widehat{\Sigma}_{S_i}$ with words in $\widehat{\Sigma}_{S_j}$ (Lemma 4.10), and reducing products of the $\widehat{v}_{ij}(V)$ (Lemma 4.11).

First, we prove a lemma relating $\widehat{v}_{ij}(V)$ and $\widehat{u}(V)$.

Lemma 4.8. *Let S_1, \dots, S_k be as in Prop. 4.7, let $1 \leq i < j \leq k$, and let $V \in \mathbb{Z}^{S_i} \otimes \mathbb{Z}^{S_j}$. Let $S \supset S_i$ and $T \supset S_j$ be disjoint sets such that $\#S \geq 2$ and $\#T \geq 3$ or vice versa. Then*

$$\widehat{\delta}(\widehat{v}_{ij}(V), \widehat{u}(V; S, T)) = O((\log \|V\|_2)^2).$$

Proof. Let x_{ab} be the (a, b) coefficient of V . There is a homotopy from $\widehat{v}_{ij}(V)$ to $\widehat{u}(V; S, T)$ going through the stages

$$\begin{aligned}
& \prod_{a \in S_i, b \in S_j} \widehat{e}_{ab}(x_{ab}) \\
& \prod_{a \in S_i, b \in S_j} \widehat{e}_{ab; S, T}(x_{ab}) && \text{by Lemma 4.4} \\
& \prod_{a \in S_i, b \in S_j} \widehat{u}(x_{ab} v_a \otimes v_b; S, T) && \text{by the definition of } \widehat{e} \\
& \widehat{u}(V; S, T) && \text{by Thm. 4.2}
\end{aligned}$$

This homotopy has area $O((\log \|V\|_2)^2)$. \square

Using this, we can prove:

Lemma 4.9. *Let S_1, \dots, S_k be as in Prop. 4.7, let $1 \leq i < j \leq k$, and let $V \in \mathbb{Z}^{S_i} \otimes \mathbb{Z}^{S_j}$. Let $w \in \widehat{\Sigma}_{S_i}^*$ be a word representing M . Then*

$$\widehat{\delta}(w \widehat{v}_{ij}(V) w^{-1}, \widehat{v}_{ij}(MV)) \leq O((\widehat{\ell}(w) + \log \|V\|_2)^2).$$

Similarly, if instead $w \in \widehat{\Sigma}_{S_j}^$ is a word representing N , then*

$$\widehat{\delta}(w \widehat{v}_{ij}(V) w^{-1}, \widehat{v}_{ij}(VN^{-1})) \leq O((\widehat{\ell}(w) + \log \|V\|_2)^2).$$

Proof. We consider the case that $w \in \widehat{\Sigma}_{S_i}^*$; the other case follows by similar methods. It suffices to prove that

$$\widehat{\delta}(w e_{ab}(t) w^{-1}, \widehat{v}_{ij}(M t z_a \otimes z_b)) = O((\widehat{\ell}(w) + \log |t|)^2),$$

and apply this inequality to each term of $\widehat{v}_{ij}(V)$ individually. Let $d \in S^c$ be such that $b \neq d$, let $T' = \{b, d\} \subset S^c$, and let $S' = (T')^c$; note that $S \subset S'$ and $\#S' \geq 3$. We can use Lemma 4.4 to replace $\lambda(w)$ by a word w' in $\Sigma_{S'}$ of length $O(\widehat{\ell}(w))$. We construct a homotopy from $w' \widehat{e}_{ab}(t) w'^{-1}$ to $\widehat{v}_{ij}(M z_a \otimes z_b)$ as follows:

$$\begin{aligned}
& w' \widehat{e}_{ab}(t) (w')^{-1} \\
& w' \widehat{u}(t z_a \otimes z_b; S', T') (w')^{-1} && \text{by Lemma 4.4} \\
& \widehat{u}(t M z_a \otimes z_b; S', T') && \text{by Lemma 4.3} \\
& \widehat{u}(t M z_a \otimes z_b; S_i, S_j) && \text{by Cor. 4.5} \\
& \widehat{v}_{ij}(M z_a \otimes z_b) && \text{by Lemma 4.8}
\end{aligned}$$

Applying this homotopy to each term in $\widehat{v}_{ij}(V)$ results in a product of terms $\widehat{v}_{ij}(V_a)$ such that $\sum V_a = MV$. We can use Lemma 4.6 to reduce this to $\widehat{v}_{ij}(MV)$. \square

Next, we fill commutators. These could be filled using Lemma 4.6.(3), but a naive application of the lemma would give a cubic filling rather than a quadratic filling.

Lemma 4.10. *Let $S, T \subset \{1, \dots, p\}$ be disjoint subsets such that $\#S, \#T \leq p-2$, and let $i, j \notin S$. Let w_S be a word in $\widehat{\Sigma}_S$ and let w_T be a word in $\widehat{\Sigma}_T$. Then*

$$\widehat{\delta}([w_S, w_T]) = O((\widehat{\ell}(w_S) + \widehat{\ell}(w_T))^2).$$

Proof. If $\#S \geq 3$ and $\#T \geq 3$, we can use Lemma 4.4 to replace w_S and w_T by words in Σ_S and Σ_T , then commute the resulting words letter by letter. Similarly, if $\#S = 1$ or $\#T = 1$, then w_S or w_T is trivial. It remains to study the case that $\#S$ or $\#T$ is 2; we can assume that $T = S^c$. Assume without loss of generality that $S = \{2, \dots, p-1\}$ and $\#T = \{1, p\}$. Since $\#S \geq 3$, we can use Lemma 4.4 to replace $\lambda(w_S)$ by a word w'_S in Σ_S at cost $O(\widehat{\ell}(w_S)^2)$.

We will first show that for any word w in Σ_S ,

$$\delta_\Gamma([w, \widehat{e}_{1p}(t)]) = O((\ell_w(w) + \log |t|)^2).$$

Let $M \in SL(S; \mathbb{Z})$ be the matrix represented by w .

We will construct a homotopy from $w\widehat{e}_{1p}(t)w^{-1}$ to $\widehat{e}_{1p}(t)$ through the curves

$$\begin{array}{ll} w[e_{12}(1), \widehat{e}_{2p}(t)]w^{-1} & \text{by Lemmas 4.4 and 4.6} \\ [we_{12}(1)w^{-1}, w\widehat{e}_{2p}(t)w^{-1}] & \text{by free insertion} \\ [\mu_{\{1\}, S}(z_1 \otimes z_2 M^{-1}), \mu_{S, p}(Mt z_2 \otimes z_p)] & \text{by Lemma 4.3} \\ \widehat{e}_{1p}(t) & \text{by Lemma 4.6} \end{array}$$

The total cost of these steps is at most $O((\widehat{\ell}(w) + \log |t|)^2)$.

The last step needs some explanation. Let $z_2 M^{-1} = (a_2, \dots, a_{p-1})$ and $Mt z_2 = (b_2, \dots, b_{p-1})$, so that we are transforming $[u, v]$ to $\widehat{e}_{1p}(t)$, where

$$u = \mu_{\{1\}, S}(z_1 \otimes z_2 M^{-1}) = \prod_{i=2}^{p-1} \widehat{e}_{1i}(a_i)$$

and

$$v = \mu_{S, \{p\}}(Mt z_2 \otimes z_p) = \prod_{i=2}^{p-1} \widehat{e}_{ip}(b_i).$$

We will move terms of u past v one by one. Each term $\widehat{e}_{1i}(a_i)$ of u commutes with every term of u and v except for $\widehat{e}_{ip}(b_i)$ and its inverse, so moving it past v only generates a single $\widehat{e}_{1p}(a_i b_i)$. This commutes with every term of u and v , so we can move it to the left side of the word. We can now cancel $\widehat{e}_{1i}(a_i)$ with its inverse. Repeating this process deletes all of the original terms of u and u^{-1} , allowing us to cancel v and v^{-1} . This leaves a product

$$\prod_{i=2}^{p-1} \widehat{e}_{1p}(a_i b_i),$$

but since $z_2 M^{-1} \cdot Mt z_2 = z_2 \cdot t z_2 = t$, where \cdot represents the dot product, we can use Lemma 4.6 to convert this to $\widehat{e}_{1p}(t)$. All of the coefficients in this process are bounded by $\|M\|_2^2 t$, and $\|M\|_2$ is exponential in $\widehat{\ell}(w)$, so this process has cost $O((\ell_w(w) + \log |t|)^2)$.

When $\ell(w)$ is large, we can get a stronger bound by breaking it into segments. Let

$$n = \left\lceil \frac{\ell_w(w)}{\log |t| + 1} \right\rceil.$$

Let $w = w_1 \dots w_n$, where the w_i are words in Σ_S of length at most $\log |t| + 1$. Then

$$\begin{aligned} \delta_\Gamma([w, \widehat{e}_{1p}(t)]) &\leq \sum_{i=1}^n \delta_\Gamma([w_i, \widehat{e}_{1p}(t)]). \\ &\leq nO((\log |t| + 1)^2) \\ &\leq O((\log |t| + 1)^2 + \ell_w(w)(\log |t| + 1)) \end{aligned}$$

The same methods show that

$$\delta_\Gamma([w, \widehat{e}_{p1}(t)]) \leq O((\log |t| + 1)^2 + \ell_w(w)(\log |t| + 1)).$$

Applying this to each term of $w_T = g_1 \dots g_k$, where $g_i \in \widehat{\Sigma}_T$, we find

$$\begin{aligned} \widehat{\delta}([w_S, w_T]) &\leq O(\widehat{\ell}(w_S)^2) + \widehat{\delta}([w'_S, w_T]) \\ &\leq O(\widehat{\ell}(w_S)^2) + \sum_{i=1}^k \widehat{\delta}([w'_S, g_i]) \\ &\leq O(\widehat{\ell}(w_S)^2) + \sum_{i=1}^k O(\widehat{\ell}(g_i)^2 + \ell_w(w'_S)\widehat{\ell}(g_i)) \\ &\leq O(\widehat{\ell}(w_S)^2) + O(k) + O(\widehat{\ell}(w_T)^2) + O(\ell_w(w'_S)\widehat{\ell}(w_T)) \\ &= O((\widehat{\ell}(w_S) + \widehat{\ell}(w_T))^2). \end{aligned}$$

□

Finally, we construct fillings of products of upper-triangular elements of $\widehat{\Sigma}$.

Lemma 4.11. *Let $w = w_1 \dots w_n$ be a word in $\widehat{\Sigma}$ representing I , where $w_i = e_{a_i, b_i}(t_i)$ for some $1 \leq a_i < b_i \leq p$. Let $h = \max\{\log |x_i|, 1\}$. If w represents the identity, then $\widehat{\delta}(w) = O(n^3 h^2)$.*

Proof. Let $N = N(\mathbb{Z})$ be the subgroup of upper-triangular integer matrices with 1's on the diagonal. Our proof is based on the seashell template in Figure 1; we describe a normal form for points in N and then describe how to fill triangles with two sides in normal form and one short side.

If

$$m = \begin{pmatrix} 1 & m_{1,2} & \dots & m_{1,p} \\ 0 & 1 & \dots & m_{2,p} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{pmatrix},$$

where $m_{ij} \in \mathbb{Z}$, then

$$\omega_0(m) = x_p(m) \dots x_1(m)$$

where

$$x_k(m) = e_{k, k+1}(m_{k, k+1}) \dots e_{k, p}(m_{k, p}).$$

This is a normal form for elements of N_P ; note that it contains at most p^2 letters.

Let $n_i = w_1 \dots w_i$. We will construct a filling of w by filling the wedges

$$\lambda(\omega_0(n_{i-1})w_i\omega_0(n_i)^{-1});$$

we can consider these fillings as homotopies from $\lambda(\omega_0(n_{i-1})w_i)$ to $\lambda(\omega_0(n_i))$.

We fill $\lambda(\omega_0(n_{i-1})w_i\omega_0(n_i)^{-1})$ by moving w_i leftward past

$$\lambda(x_{a_i-1}(n_{i-1}) \dots x_1(n_{i-1})).$$

Each letter in $x_{a_i-1}(n_{i-1}) \dots x_1(n_{i-1})$ is of the form $\widehat{e}_{ab}(x)$ with $a < a_i$, so we can move w_i leftward by repeatedly replacing subwords of the form $\widehat{e}_{ab}(x)w_i$ with $w_i\widehat{e}_{ab}(x)$ if $b \neq a_i$ (using Lemma 4.6) and with $w_i\widehat{e}_{ab_i}(xt_i)\widehat{e}_{ab}(x)$ if $b = a_i$.

Each of these steps has cost $O((\log |x| + \log |t_i|)^2)$. Since

$$\log |x| \leq \log \|n_{i-1}\|_2 = O(hn),$$

this is $O(h^2n^2)$. We repeat this process until we have moved w_i past

$$\lambda(x_{a_i-1}(n_{i-1}) \dots x_1(n_{i-1})),$$

which takes at most p^2 steps and has total cost $O(h^2n^2)$. Call the resulting word w' .

During this process, we have inserted at most p^2 additional letters, but we can partition w' into a word of the form $w' = w'_{p-1} \dots w'_1$, where w'_k is a word in the alphabet $\{e_{ki}(t)\}_{t \in \mathbb{Z}, i > k}$ representing the same element of Γ as $x_k(n_i)$. The coefficients in these words are bounded by $t_i \|n_{i-1}\|_2$. The letters in each subword commute, so we can apply Lemma 4.6 $(p + p^2)^2$ times to put the letters in each subword in order, then p^2 times to collect like terms. This reduces w' to $\omega(w_i)$. Each of these steps has cost $O(h^2n^2)$, so

$$\widehat{\delta}(\omega_0(n_{i-1})w_i\omega_0(n_i)^{-1}) = O(h^2n^2).$$

To fill w , we need to fill n such wedges, so $\delta_\Gamma(w) = O(h^2n^3)$. \square

We can use these lemmas to prove Prop. 4.7.

Proof of Prop. 4.7. Recall that w is a product of at most $3p^2 + 9p$ subwords of two types: words in $\widehat{\Sigma}_{S_i}$ or words of the form $\widehat{v}_{ij}(V)$.

Since $\#S_i \leq p - 2$ for all i , we can use Lemmas 4.9 and 4.10 to gather all of the terms in $\widehat{\Sigma}_{S_1}^*$ on the left side of the word. This takes no more than $(3p^2 + 9p)^2$ applications of the lemmas. We can do the same for all of the $\widehat{\Sigma}_{S_k}^*$. This process has cost $O(\widehat{\ell}(w)^2)$ and results in a word $p_1(w) \dots p_k(w)w'$ of length $O(\widehat{\ell}(w))$, where w' is a word of length at most $3p^2$ in the alphabet $\{e_{ab}(t)\}_{a < b, t \in \mathbb{Z}}$. Since each of the $p_i(w)$ represent the identity, so does w' . There is a c such that the coefficients in the letters of w' are bounded by $|t| \leq e^{c\widehat{\ell}(w)}$, so Lemma 4.11 provides a filling of w' with area $O(\widehat{\ell}(w)^2)$; this proves the proposition. \square

4.4. Words in $SL(p-1; \mathbb{Z}) \ltimes \mathbb{Z}^{p-1}$. The techniques of the previous section fail when $P = U(p-1, 1; \mathbb{Z})$, but the approach described in Section 3.1 allows us to decompose words in Σ_P into ω -triangles with vertices in smaller parabolic subgroups. In this section, we will construct a space on which P acts non-cocompactly and which has a quadratic Dehn function, then use fillings of words in this space to construct templates for those words.

Since P is a finite-index extension of $SL(p-1; \mathbb{Z}) \ltimes \mathbb{Z}^{p-1}$, any word in Σ_P which represents the identity can be reduced to a word in

$$\Sigma'_P := SL(p-1; \mathbb{Z}) \ltimes \mathbb{Z}^{p-1} \cap \Sigma_P$$

at cost linear in the length of the word. Druţu showed that if $p \geq 4$, then the group $H = SL(p-1; \mathbb{R}) \ltimes \mathbb{R}^{p-1}$ has a quadratic Dehn function [Dru04]. Let $\mathcal{E}_H = H/SO(p) \subset \mathcal{E}$; this fibers over $\mathcal{E}_{p-1} := SL(p-1)/SO(p-1)$ with fiber \mathbb{R}^{p-1} . If $x \in H$, let $[x]_{\mathcal{E}_H}$ be the corresponding point of \mathcal{E}_H . Our first step is to strengthen

Druţu's result by showing that a curve of length ℓ in \mathcal{E}_H can be filled by a Lipschitz map $f : D^2(\ell) \rightarrow \mathcal{E}_H$ with Lipschitz constant bounded independently of ℓ .

We prove this by using a Lipschitz analogue of Prop. 3.1; we construct a normal form for \mathcal{E}_H and then fill triangles in this normal form. First, we construct a family of curves $\widehat{\xi}(v)$ for $v \in \mathbb{R}^{p-1}$, roughly analogous to the curves $\widehat{u}(V)$ defined in Section 2.3; these correspond to elements of \mathbb{R}^{p-1} . Any element of H can be written as the product of an element of $SL(p-1; \mathbb{R})$ and an element of \mathbb{R}^{p-1} ; so we construct a normal form for H so that each curve is the concatenation of a curve in $SL(p-1; \mathbb{R})$ and a $\widehat{\xi}(v)$. To fill triangles in this normal form, we need three constructions: we need to conjugate the $\widehat{\xi}(v)$ by curves in $SL(p-1; \mathbb{R})$ (Lemma 4.15), fill triangles whose sides are the $\widehat{\xi}(v)$ (Lemma 4.16), and fill curves in $SL(p-1; \mathbb{R})$. The following remarks will help us to glue and combine Lipschitz discs.

Remark 4.12. If S is a convex polygon with non-empty interior and diameter at most 1, there is a map $S \rightarrow D^2$ whose Lipschitz constant varies continuously with the vertices of S .

Remark 4.13. If γ is a curve of length ℓ which can be filled with an Lipschitz disc, then a reparameterization of γ can be filled with a Lipschitz disc with only a small increase in the Lipschitz constant. Let $D^2(\ell) := [0, \ell] \times [0, \ell]$ as before, and let $S^1(\ell)$ be the boundary of $D^2(\ell)$, a circle of length 4ℓ . Let $\gamma : S^1(t) \rightarrow X$ be a Lipschitz map with length ℓ , and let $\gamma' : S^1(t) \rightarrow X$ be the constant-speed parameterization of γ . It is straightforward to construct a homotopy $h : S^1 \times [0, t] \rightarrow X$ between γ and γ' with Lipschitz constant $O(\text{Lip } \gamma)$, so that $h(\theta, 0) = \gamma(\theta)$ for all $\theta \in S^1$.

A filling of γ can be converted into a filling of γ and vice versa. Let $\beta : D^2(t) \rightarrow X$ be a filling of γ , and let $D' = (D^2(t) \cup S^1(t) \times [0, t]) / \sim$, where \sim identifies $D^2(t)$ with $S^1(t) \times \{0\}$. This space is bilipschitz equivalent to $D^2(t)$. Define a map $\beta : D' \rightarrow X$ so that $\beta(x) = \beta(x)$ for all $x \in D^2(t)$ and $\beta(\theta, r) = h(\theta, r)$ for all $\theta \in S^1$, $r \in [0, t]$. This is well-defined, has boundary is γ' , and its Lipschitz constant is $O(\text{Lip}(\beta) + \text{Lip}(\gamma'))$. Likewise, if β is a filling of γ' , we can use a similar construction to construct a β' such that β' fills γ and has Lipschitz constant $O(\text{Lip}(\beta) + \text{Lip}(\gamma'))$.

One application of this converts homotopies to discs; if $f_1, f_2 : [0, \ell] \rightarrow X$ are two maps with the same endpoints, and $h : [0, \ell] \times [0, \ell] \rightarrow X$ is a Lipschitz homotopy between f_1 and f_2 with endpoints fixed, then there is a disc filling $f_1 f_2^{-1}$ with Lipschitz constant $O(\text{Lip}(h))$.

Remark 4.14. For every ℓ , there is a $c(\ell)$ such that any closed curve in \mathcal{E}_H of length ℓ can be filled by a $c(\ell)$ -Lipschitz map $D^2(1) \rightarrow \mathcal{E}_H$. This follows from compactness and the homogeneity of \mathcal{E}_H .

We define a family of curves $\widehat{\xi}(v)$ which connect I to $v \in \mathbb{R}^{p-1} \subset H$. If $v = 0$, let $\widehat{\xi}(v)$ be constant. If $v \in \mathbb{R}^{p-1}$, $v \neq 0$, let $\kappa = \max\{\|v\|_2, 1\}$ and let

$$n(v) := \frac{v}{\kappa},$$

so that $\|n(v)\| \leq 1$. Let $v_1 = v/\|v\|_2, v_2, \dots, v_{p-1} \in \mathbb{R}^{p-1}$ be an orthonormal basis of \mathbb{R}^{p-1} . Let $D(v)$ be the matrix taking

$$\begin{aligned} v_1 &\mapsto \kappa v_1 \\ v_i &\mapsto \kappa^{-1/(p-2)} v_i. \end{aligned}$$

and let $\mathcal{D}(v)$ be the curve $t \mapsto D(v)^t$ for $0 \leq t \leq 1$. Then $D(v)u(n(v))D(v)^{-1} = u(v)$, and we can let $\widehat{\xi}(v)$ be $\mathcal{D}(v)u(n(v))\mathcal{D}(v)^{-1}$. This has length $O(\overline{\log \|v\|})$, where $\overline{\log} x = \max\{1, \log x\}$.

We can prove the following analogue of Lemma 4.3 for the $\widehat{\xi}(V)$'s:

Lemma 4.15. *Let $\gamma : [0, 1] \rightarrow SL(p-1)$ be a curve connecting I and M and let $v \in \mathbb{R}^{p-1}$. Let*

$$w = \gamma \widehat{\xi}(v) \gamma^{-1} \widehat{\xi}(Mv)^{-1}.$$

There is a map $f : D^2(\ell_c(w)) \rightarrow \mathcal{E}_H$ such that $f|_{\partial D^2}$ is $[w]_{\mathcal{E}_H}$ and f has Lipschitz constant bounded independently of w .

Proof. If $v = 0$, then $w = \gamma \gamma^{-1}$, so we may assume that $v \neq 0$. If $\ell_c(w) \leq 1$, we can use Remark 4.14 to fill w , so we also assume that $\ell_c(w) \geq 1$.

To simplify the notation, let $D(v)$ represent the curve $\mathcal{D}(v)$.

By the definition of $\widehat{\xi}$,

$$w = \gamma \mathcal{D}(v) u(n(v)) \mathcal{D}(v)^{-1} \gamma^{-1} \mathcal{D}(Mv) u(-n(Mv)) \mathcal{D}(Mv)^{-1}.$$

Let

$$\gamma' = \mathcal{D}(Mv)^{-1} \gamma \mathcal{D}(v).$$

Changing the basepoint of w , we obtain the curve

$$w_1 = \gamma' u(n(v)) (\gamma')^{-1} u(-n(Mv)).$$

This can be filled as shown in Figure 4, but the horizontal lines in the figure may be exponentially large. The color-coding in the diagrams corresponds to direction; blue lines represent curves in $SL(p-1)$ and red lines represent curves $\widehat{\xi}(v)$. We will use a homotopy in $SL(p-1)$ to replace w_1 with a “thin rectangle” in which the horizontal lines are short.

We will construct a map $[0, 2\ell_c(w) + 1] \times [\ell_c(w)] \rightarrow \mathcal{E}_H$ whose boundary is a parameterization of w . The domain of this map is divided into two $\ell_c(w) \times \ell_c(w)$ squares and a $\ell_c(w) \times 1$ rectangle (Fig. 5); the squares will correspond to the homotopy in $SL(p-1)$ mentioned above, and bound a central thin rectangle. We will map the boundaries of each of these shapes into \mathcal{E}_H by Lipschitz maps and construct Lipschitz discs in \mathcal{E}_H with those boundaries. Indeed, each of the edges of the figure are marked by curves, and we can map each edge into \mathcal{E}_H as the constant-speed parameterization of the corresponding curve. Our bounds on the lengths of these curves will ensure that these maps are Lipschitz.

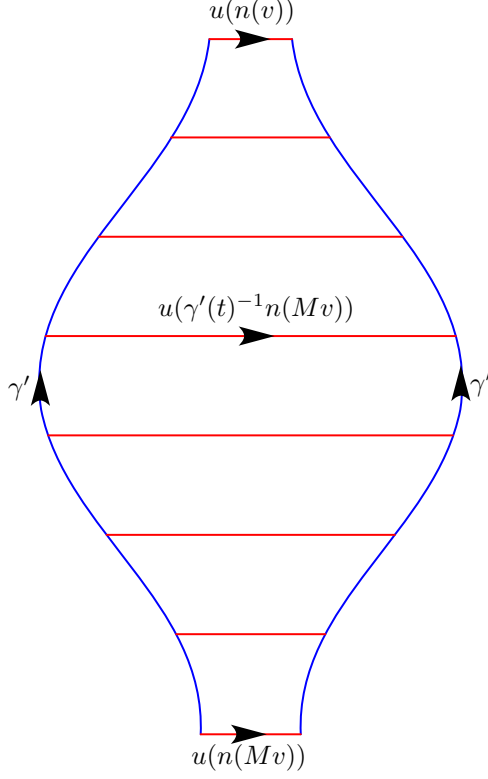
Let

$$S := \{m \mid m \in SL(p-1), \|m^{-1}n(Mv)\|_2 \leq 1\},$$

and let

$$M' := D(Mv)^{-1} M D(v)$$

be the endpoint of γ' . Since $n(v) = (M')^{-1}n(Mv)$, the endpoints of γ' both lie in S . We will construct a curve σ in S which has the same endpoints as γ' and has length comparable to $\ell_c(\gamma')$. First, let $K \in SO(p-1)$ be a matrix such that $K^{-1}Mv/\|Mv\|_2 = v/\|v\|$. There is a curve σ_K in $SO(p-1) \subset S$ connecting I to

FIGURE 4. An exponential filling of $\gamma'u(n(v))(\gamma')^{-1}u(-n(Mv))$

K . The vectors $Kn(v)$ and $n(Mv)$ are both multiples of Mv , so we can let $\lambda > 0$ be such that $Kn(v) = \lambda n(Mv)$. Note that $|\log \lambda| = O(\ell_c(w))$.

Let $D \in SL(p-1)$ be the matrix such that $D^{-1}v = \lambda v$ and $D^{-1}w = \lambda^{-1/(p-1)}w$ when w is perpendicular to v , so that

$$D^{-1}K^{-1}n(Mv) = n(v).$$

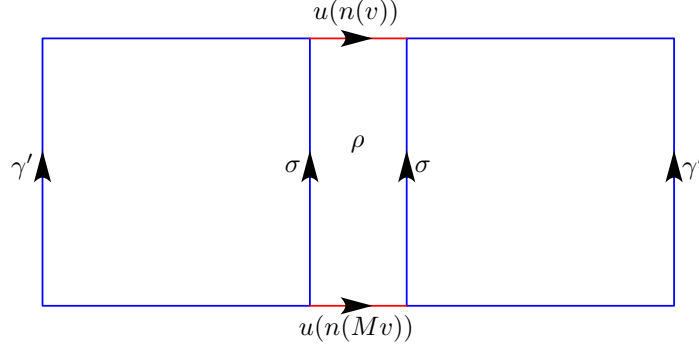
Note that $\log \|D\|_2 = O(|\log \lambda|) = O(\ell_c(w))$. Let σ_D be the path $t \mapsto D^t$, $t \in [0, 1]$. This connects I and D , and the concatenation $\sigma_K \sigma_D$ connects I to KD in S . We next connect KD and M' inside S . Note that if $SL(p-1)_v$ is the stabilizer of v , then $KD \cdot SL(p-1)_v \subset S$, and $(KD)^{-1}M'n(v) = n(v)$, so $(KD)^{-1}M' \in SL(p-1)_v$. The set $SL(p-1)_v$ is isometric to $SL(p-2) \ltimes \mathbb{R}^{p-2}$, and since $p \geq 4$, it is quasi-isometrically embedded in H , and there is a path σ_0 in $SL(p-1)_v$ connecting I to $(KD)^{-1}M'$, with length $O(\log \|(KD)^{-1}M'\|_2) = O(\ell_c(w))$. The path

$$\sigma = \sigma_K \sigma_D \sigma_0$$

is then contained in S and connects I to M' . Adding the lengths of the components, we find $\ell_c(\sigma) = O(\ell_c(w))$. Since $\ell_c(w) \geq 1$ and the edges marked σ in Fig. 5 have length $\ell_c(w)$, the map on those edges is Lipschitz for some constant independent of w .

The boundaries of the shapes in the figure are thus $[\sigma\gamma^{-1}]_{\mathcal{E}_H}$ and

$$w_2 = [\sigma u(n(v))\sigma^{-1}u(-n(Mv))]_{\mathcal{E}_H}.$$

FIGURE 5. A quadratic filling of $\gamma' u(n(v)) (\gamma')^{-1} u(-n(Mv))$

The first curve, $\sigma\gamma^{-1}$, is a closed curve in $SL(p-1)$ of length $O(\ell_c(w))$. Since $SL(p-1)/SO(p-1)$ is non-positively curved, this curve has a filling in \mathcal{E}_H with area $O(\ell_c(w)^2)$. This can be taken to be a c -Lipschitz map from $D^2(\ell_c(w))$, where c depends only on p .

The second curve is the boundary of a “thin rectangle”. That is, there is a Lipschitz map

$$\begin{aligned} \rho : [0, \ell_c(w)] \times [0, 1] &\rightarrow H \\ \rho(x, t) &= \sigma(x) u(t\sigma(x)^{-1} n(Mv)) = u(tn(Mv)) \sigma(x) \end{aligned}$$

which sends the four sides of the rectangle to $\sigma, u(n(v)), \sigma^{-1}$, and $u(-n(Mv))$. Projecting this disc to \mathcal{E}_H gives a Lipschitz filling of w_2 .

We glue these discs together to get a Lipschitz map from the rectangle to \mathcal{E}_H . The boundary of the rectangle is a Lipschitz reparameterization of $[w]_{\mathcal{E}_H}$, so we can use Rem. 4.13 to get a filling of $[w]_{\mathcal{E}_H}$. \square

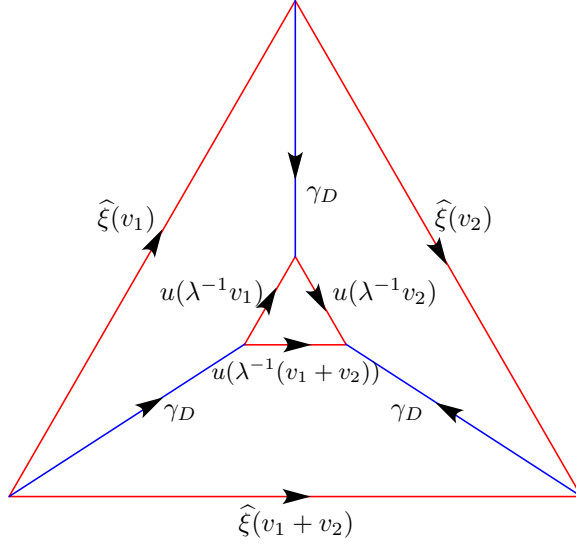
This lets us fill many curves in H .

Lemma 4.16. *Let $v_1, v_2 \in \mathbb{R}^{p-1}$. If $w = \widehat{\xi}(v_1) \widehat{\xi}(v_2) \widehat{\xi}(v_1 + v_2)^{-1}$, there is a map $f : D^2(\ell_c(w)) \rightarrow \mathcal{E}_H$ such that $f|_{\partial D^2} = [w]_{\mathcal{E}_H}$ and $\text{Lip}(f)$ is bounded independently of w .*

Proof. As before, we may assume that $\ell_c(w) > 3$. Let $S = \langle v_1, v_2 \rangle \subset \mathbb{R}^{p-1}$ be the subspace generated by the v_i and let $\lambda = \max\{\|v_1\|_2, \|v_2\|_2, \|v_1 + v_2\|_2\}$. Let $D \in SL(p-1)$ be the matrix such that $Ds = \lambda s$ for $s \in S$ and $Dt = \lambda^{-1/(p-1-\dim(S))} t$ for vectors t which are perpendicular to S ; this is possible because $\dim(S) \leq 2$ and $p \geq 4$.

Let γ_D be the curve $t \mapsto D^t$ for $0 \leq t \leq 1$; this has length $O(\log \lambda) = O(\ell_c(w))$. We construct a filling of $[w]_{\mathcal{E}_H}$ based on a triangle with side length $\ell_c(w)$ as in Figure 6. The central triangle in the figure has side length 1; since $\ell_c(w) \geq 3$, the trapezoids around the outside are bilipschitz equivalent to discs $D^2(\ell)$ with Lipschitz constant bounded independently of w . Let f take each edge to H as labeled, and give each edge a constant-speed parameterization; f is Lipschitz on each edge, with a Lipschitz constant independent of the v_i . Let \bar{f} be the projection of f to \mathcal{E}_H . We’ve defined \bar{f} on the edges in the figure; it remains to extend it to the interior of each cell.

The map \bar{f} sends the boundary of the center triangle to a curve of length at most 3, so we can use Rem. 4.14 to extend \bar{f} to its interior. The map \bar{f} sends the

FIGURE 6. A quadratic filling of $\hat{\xi}(v_1)\hat{\xi}(v_2)\hat{\xi}(v_1 + v_2)^{-1}$

boundary of each trapezoid to a curve of the form

$$(12) \quad [\hat{\xi}(v_i)^{-1}\gamma_D u(\lambda v_i)\gamma_D^{-1}]_{\mathcal{E}_H}.$$

Lemma 4.15 gives Lipschitz discs filling such curves. Each of these fillings has Lipschitz constant bounded independently of w , so the resulting map on the triangle also has Lipschitz constant bounded independently of w . \square

The group H has a normal form based on the semidirect product structure. Let $g \in H$. Then $g = Mu(v)$ for some $M \in SL(p-1)$ and $v \in \mathbb{R}^{p-1}$. Let γ_M be a geodesic connecting I to M and define $\omega_H(g) = \gamma_M \hat{\xi}(v)$. Then $\ell_c(\omega_H(g)) = O(\log \|g\|_2)$.

We can use the previous lemmas to fill ω_H -triangles.

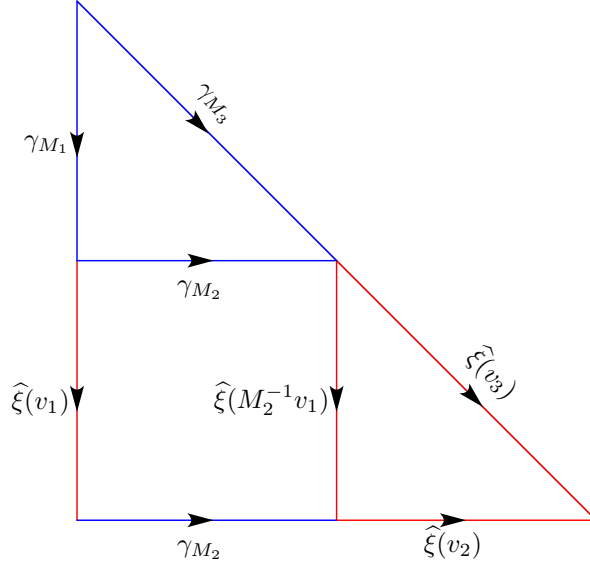
Lemma 4.17. *If $g_1, g_2 \in H$ and*

$$w = \omega_H(g_1)\omega_H(g_2)\omega_H(g_1g_2)^{-1},$$

then there is a map $f : D^2(\ell_c(w)) \rightarrow \mathcal{E}_H$ such that $f|_{\partial D^2} = [w]_{\mathcal{E}_H}$ and $\text{Lip}(f) \leq c$.

Proof. As before, assume that $\ell_c(w) \geq 1$. Let $g_3 = g_1g_2$ and let $M_i \in SL(p-1)$ and $v_i \in \mathbb{R}^{p-1}$ be such that $g_i = M_i u(v_i)$ for $i = 1, 2, 3$. We construct a filling of w as in Figure 7, which depicts an isosceles right triangle with legs of length $2\ell_c(w)$, divided into a square and two congruent triangles; this is bilipschitz equivalent to $D^2(\ell_c(w))$, and we call it Δ . Let $f : \Delta^{(1)} \rightarrow H$ be the map from the 1-skeleton of the figure to H defined by sending each edge to its corresponding curve, parameterized with constant speed, and let $\bar{f} = [f]_{\mathcal{E}_H}$. This is a Lipschitz map on the 1-skeleton of the triangle, with Lipschitz constant independent of w .

The map \bar{f} sends the square in the figure to a curve which can be filled using Lemma 4.15, sends the blue (upper left) triangle to a curve in \mathcal{E}_{p-1} which can be filled using non-positive curvature, and sends the red triangle (lower right) to a curve which can be filled using Lemma 4.16. Combining these fillings proves the lemma. \square

FIGURE 7. A quadratic filling of $\omega_H(g_1)\omega_H(g_2)\omega_H(g_3)^{-1}$

We can construct Lipschitz fillings of arbitrary curves from these triangles (cf. [Gro96] and Prop. 3.1).

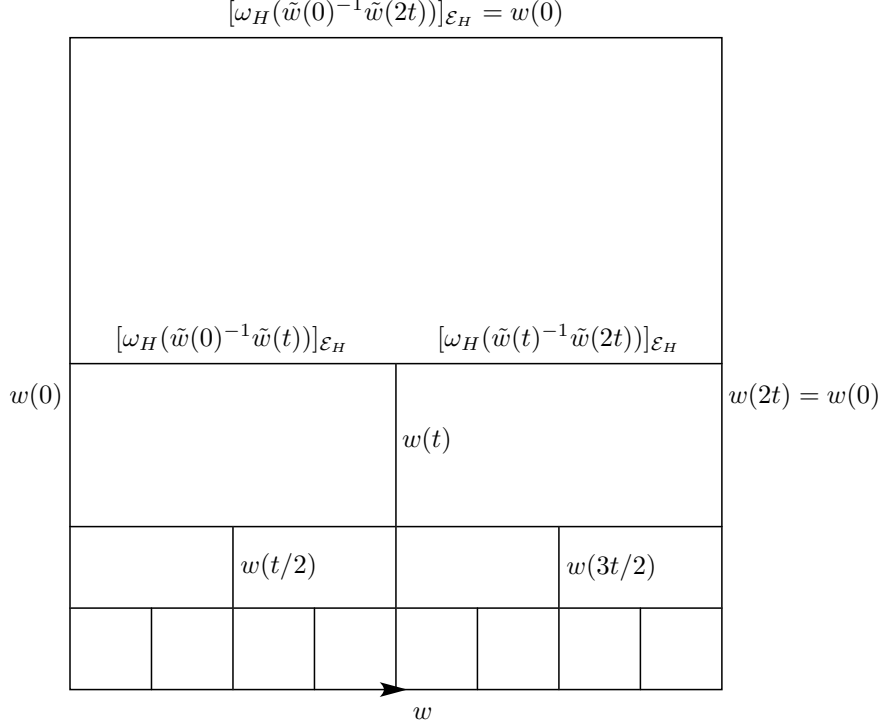
Proposition 4.18. *If w is a closed curve in \mathcal{E}_H , there is a filling $f : D^2(\ell_c(w)) \rightarrow H$ of w such that $\text{Lip}(f)$ is bounded independently of w .*

Proof. The proof is based on Gromov's construction of Lipschitz extensions in [Gro96]. Assume that $\ell_c(w) > 1$; otherwise, the lemma follows from Rem. 4.14. Let $k = \lceil \log \ell_c(w) \rceil$, let $t = 2^k$, and parameterize w as a map $w : [0, 2t] \rightarrow \mathcal{E}_H$ with constant speed. Let $\tilde{w} : [0, 2t] \rightarrow H$ be a map (not necessarily continuous) such that $[\tilde{w}(x)]_{\mathcal{E}_H} = w(x)$ for all x . We construct $f : D^2(2t) \rightarrow \mathcal{E}_H$ as in Figure 8. The figure depicts $k + 1$ rows of rectangles; the top row has one $2^k \times 2^{k+1}$ rectangle, while the i th from the top consists of 2^{i-1} rectangles of dimensions $2^{k-i+1} \times 2^{k-i}$. The bottom row is an exception, consisting of t squares of side length 1. Note that a cell in the i th row is bilipschitz equivalent to $D^2(2^{k-i})$. Call the resulting complex X .

We label all the edges of X by curves in \mathcal{E}_H . First, we label all the vertical edges into \mathcal{E}_H as constant curves; the vertical edges with x -coordinate a are labeled by $w(a)$. We label horizontal edges using the normal form; the edge from (x_1, y) to (x_2, y) is labeled $[\omega_H(\tilde{w}(x_1)^{-1}\tilde{w}(x_2))]_{\mathcal{E}_H}$, except for the bottom edge of X , which is labeled w . We construct a map $f : X^{(1)} \rightarrow \mathcal{E}_H$ by sending each edge to the constant-speed parameterization of its label; this map is Lipschitz, with Lipschitz constant independent of w , and the map sends the boundary of X to a Lipschitz reparameterization of w .

If Δ is a 2-cell from the i th row of the figure, $i \leq k$, then f sends its boundary to

$$[\omega_H(g_1)\omega_H(g_2)\omega_H(g_1g_2)^{-1}]_{\mathcal{E}_H},$$

FIGURE 8. A filling of w in \mathcal{E}_H

where

$$g_1 = \tilde{w}(n2^{-i})^{-1}\tilde{w}((n+1)2^{-i})$$

$$g_2 = \tilde{w}((n+1)2^{-i})^{-1}\tilde{w}((n+2)2^{-i}).$$

These are the triangles filled in Lemma 4.17, so we can extend the map on the boundary of the cell to a map on the whole cell. Furthermore, the Lipschitz constants of these maps are uniformly bounded.

This defines a map on all of the rows except the bottom row. Cells in the bottom row have boundaries of the form

$$w|_{[n,n+1]}[\omega_H(w(n)^{-1}w(n+1)^{-1})^{-1}]\mathcal{E}_H$$

for some n ; this has bounded length, so we can fill each of the squares in the bottom row by maps with uniformly bounded Lipschitz constants. \square

At this point, we are in a similar situation to the one in Section 3.1; if $w \in \Sigma_H^*$, the lemma gives a filling of the curve corresponding to w , but this filling may travel far from $H(\mathbb{Z}) = U(p-1, 1; \mathbb{Z})$. The next step is to replace this filling by a filling in $H(\mathbb{Z})$.

First, we will need notation like that in Sec. 2.4 for sets and maps associated to $SL(p-1; \mathbb{R})$. Let \mathcal{E}_{p-1} be the symmetric space $SL(p-1; \mathbb{R})/SO(p-1)$, let $\mathcal{S}_{p-1} \subset \mathcal{E}_{p-1}$ be a Siegel set, let N_{p-1}^+ and A_{p-1}^+ be the sets of unipotent and diagonal matrices used to define \mathcal{S}_{p-1} , let ρ_{p-1} be the map $\mathcal{E}_{p-1} \rightarrow SL(p-1; \mathbb{Z})$ defined using \mathcal{S}_{p-1} , and let $\phi_{p-1} : \mathcal{E}_{p-1} \rightarrow A_{p-1}^+$ be the analogue of the original ϕ .

Let

$$N_H^+ := \{u(v) \mid v \in [-1/2, 1/2]^{p-1} \subset \mathbb{R}^{p-1}\} \subset H.$$

Recall that

$$\mathcal{S}_{p-1} = [N_{p-1}^+ A_{p-1}^+]_{\mathcal{E}_{p-1}};$$

we define a fundamental set for the action of $H(\mathbb{Z})$ on \mathcal{E}_H by

$$\mathcal{S}_H = [N_H^+ N_{p-1}^+ A_{p-1}^+]_{\mathcal{E}_H}.$$

There is a projection $\mu : H \rightarrow SL(p-1)$ which descends to a map $\mu_{\mathcal{E}} : \mathcal{E}_H \rightarrow \mathcal{E}_{p-1}$; note that $\mu_{\mathcal{E}}(\mathcal{S}_H) = \mathcal{S}_{p-1}$. This lets us define a $\rho_H : \mathcal{E}_H \rightarrow H(\mathbb{Z})$ which is compatible with ρ_{p-1} , so that $x \in \rho_H(x)\mathcal{S}_H$ and $\mu(\rho_H(x)) = \rho_{p-1}(\mu_{\mathcal{E}}(x))$ for all x .

All of the results in Sec. 2.4 apply in this context, possibly with different constants.

We will prove the following (cf. Prop. 3.2):

Lemma 4.19. *There is a c' such that if $w = w_1 \dots w_{\ell}$ is a word in Σ_H , there is a template for w such that*

- *If $g_1, g_2, g_3 \in \Gamma$ are the labels of a triangle in the template, then either $d_{\Gamma}(g_i, g_j) \leq 2c'$ for all i, j or there is a k such that $g_i^{-1}g_j \in U(k, p-1-k, 1)$ for all i, j .*
- *τ has $O(\ell^2)$ triangles, and if the i th triangle of τ has vertices labeled $(g_{i,1}, g_{i,2}, g_{i,3})$, then*

$$\sum_i (d_{\Gamma}(g_{i,1}, g_{i,2}) + d_{\Gamma}(g_{i,1}, g_{i,3}) + d_{\Gamma}(g_{i,2}, g_{i,3}))^2 = O(\ell^2).$$

Similarly, if the i th edge of τ has vertices labeled $(h_{i,1}, h_{i,2})$, then

$$\sum_i d_{\Gamma}(h_{i,1}, h_{i,2})^2 = O(\ell^2).$$

Proof. As in Sec. 2.1, we can choose curves in \mathcal{E}_H which correspond to generators of $H(\mathbb{Z})$ and concatenate them into a curve $\bar{w} : [0, \ell] \rightarrow \mathcal{E}_H$ such that $\text{Lip}(\bar{w})$ is bounded independently of w and $\bar{w}(i) = [w_1 \dots w_i]_{\mathcal{E}_H}$ for $i = 0, 1, \dots, \ell$. By Prop. 4.18, we can construct a filling $f : D^2(\ell) \rightarrow \mathcal{E}_H$ such that $f|_{\partial D^2} = \bar{w}$ and $\text{Lip}(f)$ is bounded independently of w . Let $t = 2^{\lceil \log_2 \ell \rceil}$. By Remarks 4.12 and 4.13, we can reparameterize f to get a map $f' : D^2(t) \rightarrow \mathcal{E}_H$ so that $f'(x, 0) = \bar{w}(x)$ for $x < t$ and $f' = [I]_{\mathcal{E}_H}$ on the rest of the boundary of $D^2(t)$. Note that $\text{Lip}(f')$ is also bounded independently of w , say by c_f . As in Sec. 3.1, we will use this disc to build a template.

Let $r_0 : \mathcal{E}_{p-1} \rightarrow \mathbb{R}$ be given by

$$r_0(x) = \frac{d_{\mathcal{M}_{p-1}}([x]_{\mathcal{M}_{p-1}}, [I]_{\mathcal{M}_{p-1}})}{2(p-1)^2} - c.$$

where c is as in Cor. 2.11, i.e., if $x, y \in \mathcal{E}_{p-1}$ satisfy

$$d_{\mathcal{E}_{p-1}}(x, y) < r_0(x)$$

and $g, h \in SL(p-1; \mathbb{Z})$ satisfy $x \in g\mathcal{S}_{p-1}$ and $y \in h\mathcal{S}_{p-1}$, then there is a $1 \leq j < p-1$ depending only on x such that $g^{-1}h \in U(j, p-1-j)$.

Note also that if $g \in H(\mathbb{Z})$ and $x \in g\mathcal{S}_H$, then $x = [gnn'a]_{\mathcal{E}}$ for some $n \in N_H^+$, $n' \in N_{p-1}^+$, and $a \in A_{p-1}^+$. Then since N_H^+ and N_{p-1}^+ are bounded,

$$d_{\mathcal{E}}(x, [g]_{\mathcal{E}}) = d_{A_{p-1}}(I, a) + O(1),$$

and, by Lem. 2.8

$$d_{A_{p-1}}(I, a) \leq d_{\mathcal{M}_{p-1}}([I]_{\mathcal{M}_{p-1}}, \mu_{\mathcal{E}}(x)) + c''$$

for some $c'' > 0$, independently of x . We thus have

$$(13) \quad d_{\mathcal{E}}(x, [g]_{\mathcal{E}}) = O(r_0(\mu(x))).$$

Let $r : D^2(t) \rightarrow \mathbb{R}^+$ be

$$r(v) = \max\{1, \frac{r_0(\mu_{\mathcal{E}}(f'(v)))}{2c_f}\}.$$

This is 1-Lipschitz. Let τ be the adaptive triangulation τ_r constructed in Prop. 3.3. If v is an interior vertex of τ , label it $\rho_H(f'(v))$, and label each boundary vertex by the corresponding $w(i)$ or by I . In either case, if v is labeled by $g \in H(\mathbb{Z})$, then $f'(v) \in g\mathcal{S}_H$ and $\mu_{\mathcal{E}}(f'(v)) \in \mu(g)\mathcal{S}_{p-1}$. As in the proof of Prop. 3.2, each lattice point on the boundary of $D^2(t)$ is a vertex of τ , so τ is a template for w .

Let (a_1, a_2) be an edge of τ , and say that a_i is labeled by $g_i \in H(\mathbb{Z})$. Decompose each g_i as $h_i u(v_i)$ for some $h_i \in SL(p-1; \mathbb{Z})$, $v_i \in \mathbb{Z}^{p-1}$. Prop. 3.3.(3) and the fact that f' is c_f -Lipschitz ensure that

$$d_{\mathcal{E}_H}(f'(a_1), f'(a_2)) \leq 2c_f r(a_1) \leq \max\{2c_f, r_0(\mu_{\mathcal{E}}(f'(a_1)))\}.$$

If $2c_f \leq r_0(\mu_{\mathcal{E}}(f'(a_1)))$ then

$$d_{\mathcal{E}_{p-1}}(\mu(f'(a_1)), \mu(f'(a_2))) \leq r_0(\mu_{\mathcal{E}}(f'(a_1))),$$

and by the choice of c , there is a j such that $h_1^{-1}h_2 \in U(j, p-1-j) \subset SL(p-1; \mathbb{Z})$ and thus $g_1^{-1}g_2 \in U(j, p-1-j, 1)$.

On the other hand, if

$$r_0(\mu_{\mathcal{E}}(f'(a_1))) \leq 2c_f,$$

then $d_{\mathcal{E}_{p-1}}(\mu_{\mathcal{E}}(f'(a_1)), \mu_{\mathcal{E}}(f'(a_2))) \leq 2c_f$. Furthermore, $f'(a_1)$ and $f'(a_2)$ are a bounded distance from $H(\mathbb{Z})$, so g_1 and g_2 are a bounded distance apart in Γ . This proves the first part of the lemma.

To prove the second part, we use Thm. 2.3. If (v_1, v_2) is an edge of τ and v_i is labeled by g_i , we know that $r(v_i) = O(d_{D^2}(v_1, v_2))$. By (13),

$$d_{\mathcal{E}}([g_i]_{\mathcal{E}}, f(v_i)) = O(d_{D^2}(v_1, v_2)),$$

so

$$\begin{aligned} d_{\mathcal{E}}([g_1]_{\mathcal{E}}, [g_2]_{\mathcal{E}}) &\leq d_{\mathcal{E}}([g_1]_{\mathcal{E}}, f(v_1)) + d_{\mathcal{E}}(f(v_1), f(v_2)) + d_{\mathcal{E}}(f(v_2), [g_2]_{\mathcal{E}}) \\ &= O(d_{D^2}(v_1, v_2)). \end{aligned}$$

By Thm. 2.3, this implies

$$d_{\Gamma}(g_1, g_2) = O(d_{D^2}(v_1, v_2))$$

as well. The second part of the lemma then follows from the bounds in Prop. 3.3. \square

This reduces the problem of filling words in H to the problem of filling ω -triangles with vertices in subgroups of the form $U(k, p-1-k, 1)$. These subgroups can be handled by the methods of the previous subsection, reducing such an ω -triangle to words in $\hat{\Sigma}_k$ and $\hat{\Sigma}_{p-1-k}$. These will be filled in the next section.

4.5. Filling words in $\widehat{\Sigma}_{S_i}$. In the last sections, we reduced the problem of filling an ω -triangle with vertices in a parabolic subgroup to the problem of filling a word w in $\widehat{\Sigma}_q$ for $q \leq p-2$. When $q \geq 3$, this is straightforward; we can use Lemma 4.4 to replace w by a word w' in Σ_q and use Prop. 3.2 to reduce the problem of filling w' to the problem of filling ω -triangles in a smaller parabolic subgroup.

When $q = 2$, this method isn't possible; in this section, we will describe how to fill a word w in $\widehat{\Sigma}_2$. We will use the notation of Sec. 2.4; let $\mathcal{E}_2 := SL(2)/SO(2)$ and $\mathcal{M}_2 := SL(2; \mathbb{Z}) \backslash \mathcal{E}_2$, and let \mathcal{S}_2 be a Siegel set. Let $\phi : \mathcal{E}_2 \rightarrow A_2^+$ and $\phi_i : \mathcal{E}_2 \rightarrow \mathbb{R}$, $i = 1, 2$ be as in Sec. 2.4

Lemma 4.20. *If w is a word in $\widehat{\Sigma}_2$, then*

$$\widehat{\delta}(w) \leq O(\widehat{\ell}(w)^2).$$

Proof. As in the proof for the whole group, we proceed by constructing a template for w using Prop. 3.3 and then filling the template. The largest change from Sec. 3.1 is that the curve α will not be in the thick part of \mathcal{E}_2 .

Let $w = w_1 \dots w_n$, where $w_i \in \widehat{\Sigma}_2$. The first thing to do is to construct a curve α in \mathcal{E} which corresponds to w . First, note that we can use Lemma 4.6 to replace all occurrences of $e_{21}(x)$ in w by $ge_{12}(-x)g^{-1}$, where g is a word representing a Weyl group element. This has cost $O(\widehat{\ell}(w)^2)$, so we may assume that $e_{21}(x)$ does not occur in w for $|x| \geq 1$.

Let $w(i) = w_1 \dots w_i \in \Gamma_2$, and let $\ell_i = \widehat{\ell}(w_1 \dots w_i)$ for $i = 0, \dots, n$. Note that ℓ_i is always an integer. We will construct a curve $\alpha : [0, \widehat{\ell}(w)] \rightarrow \mathcal{E}_q$ which has speed bounded independently of w and has the property that if $t \in \mathbb{Z}$, $\alpha(t) \in w(\eta(t))\mathcal{S}_2$, where η is a non-decreasing function such that $\eta(\ell_i) = i$.

The curve will be a concatenation of curves $\alpha_i : [0, \widehat{\ell}(w_i)] \rightarrow \mathcal{E}_2$ corresponding to w_i . If $\widehat{\ell}(w_i) < 3$, let α_i be the geodesic connecting $[w(i)]_{\mathcal{E}_2}$ and $[w(i+1)]_{\mathcal{E}_2}$ on the interval $[0, 1]$, and let it take the constant value $[w(i+1)]_{\mathcal{E}_2}$ on $[1, \widehat{\ell}(w_i)]$. Then $\alpha_i(t) \in w(i)\mathcal{S}_q$ if $t = 0$ and $\alpha_i(t) \in w(i+1)\mathcal{S}_q$ for $t \geq 1$.

If $\widehat{\ell}(w_i) \geq 3$, then let x be such that $w_i = e_{12}(x)$. Let

$$D = \text{diag}(e, e^{-1})$$

and note that $D^x \in \mathcal{S}_2$ for all $x \geq 0$. Let $t_1 \in \mathbb{Z}$ be such that

$$\frac{\widehat{\ell}(w_i)}{3} \leq t_1 < t_1 + 1 \leq \frac{2\widehat{\ell}(w_i)}{3}.$$

Let $g : [0, \widehat{\ell}(w_i)] \rightarrow SL(2; \mathbb{R})$ be the concatenation of geodesic segments connecting

$$\begin{aligned} p_1 &= I \\ p_2 &= D^{\log(|x|)/2} \\ p_3 &= D^{\log(|x|)/2} e_{12}(\pm 1) \\ p_4 &= D^{\log(|x|)/2} e_{12}(\pm 1) D^{-\log(|x|)/2} = e_{12}(x) = w_i. \end{aligned}$$

Here the sign of ± 1 is the same as the sign of x . Parameterize this curve so that $g|_{[0, t_1]}$ connects p_1 and p_2 , $g|_{[t_1, t_1+1]}$ connects p_2 and p_3 , and $g|_{[t_1+1, \widehat{\ell}(w_i)]}$ connects p_3 and p_4 . This curve has velocity bounded independently of x . Furthermore, if $t \in \mathbb{Z}$, then $g(t) \in \mathcal{S}_2$ if $t \leq t_1$ and $g(t) \in w_i\mathcal{S}$ if $t \geq t_1 + 1$. Let $\alpha_i(t) = [\gamma_i g(t)]_{\mathcal{E}_2}$.

Let $\alpha : [0, \widehat{\ell}(w)] \rightarrow \mathcal{E}_2$ be the concatenation of the α_i . From here, we largely follow the proof of Prop. 3.2; we construct a filling f of α , an adaptive triangulation τ , and a template based on τ so that a vertex x of τ is labeled by an element γ such that $f(x) \in \gamma\mathcal{S}_2$.

Let d be the smallest power of 2 larger than $\widehat{\ell}(w)$ and let $\alpha' : [0, d] \rightarrow \mathcal{E}_2$ be the extension of α to $[0, d]$, where $\alpha'(t) = [I]_{\mathcal{E}_2}$ when $t \geq \widehat{\ell}(w)$. As in Sec. 3.1, let $\gamma_{x,y} : [0, 1] \rightarrow \mathcal{E}$ be the geodesic from x to y and define a homotopy $f : [0, d] \times [0, d] \rightarrow \mathcal{E}$ by

$$f(x, y) = \gamma_{\alpha'(x), \alpha'(0)}(y/d).$$

This is Lipschitz with a constant c independent of w and has area $O(\widehat{\ell}(w)^2)$.

As in the proof of Prop. 3.2, let $r_0 : \mathcal{E} \rightarrow \mathbb{R}$ be

$$r_0(x) = \frac{d_{\mathcal{M}_2}([x]_{\mathcal{M}_2}, [I]_{\mathcal{M}_2})}{2} - c'$$

and let $r : D^2(d) \rightarrow \mathbb{R}$ be

$$r(v) = \max\{1, \frac{r_0(x)}{4c}\}.$$

As before, we use this function to construct a triangulation τ_r of $D^2(d)$. One major difference is that r is not necessarily small on the boundary of $D^2(d)$; on the other hand, $\alpha(\ell_i) = [w(i)]_{\mathcal{E}_2}$, so each point $(\ell_i, 0)$ is a vertex of τ_r . Label the boundary vertices of τ_r so that $(t, 0)$ is labeled by $w(\eta(t))$ and all the others are labeled I . Label the interior vertices so that v is labeled by $\rho(f(v))$. If v is a vertex, let $g_v \in \Gamma$ be its label. As in the earlier construction, $f(v) \in g_v\mathcal{E}_2$ for all v . Furthermore, the set of boundary labels is exactly $\{w(0), \dots, w(n)\}$ and, since $\omega(g, g)$ is the empty word for all g , the boundary word of τ is

$$w_\tau = \omega(w_1) \dots \omega(w_n);$$

since $\omega(e_{12}(t)) = \widehat{e}_{12}(t)$ and all the other letters in w have bounded length, we have

$$\delta_\Gamma(\lambda(w), w_\tau) = O(\widehat{\ell}(w)).$$

A filling of the triangles in τ thus gives a filling of w . As in the earlier construction, each triangle of τ either has short edges and thus a bounded filling area, or has vertices whose labels lie in a translate of a parabolic subgroup. In this case, that parabolic subgroup must be $U(1, 1; \mathbb{Z})$, and Lemma 4.6 allows us to fill any such triangle with quadratic area. Prop. 3.3.(4) thus implies that $\widehat{\delta}(w) = O(\widehat{\ell}(w)^2)$, as desired. \square

4.6. Proof of Thm. 1.2. Let $q < p$. We claim that if w is a word in Σ_q which represents the identity, then

$$(14) \quad \delta_\Gamma(w) = O(\ell(w)^2),$$

and that if w is a word in $\widehat{\Sigma}_q$ which represents the identity, then

$$(15) \quad \widehat{\delta}(w) = O(\widehat{\ell}(w)^2).$$

We proceed by induction on q .

If $q = 2$, then (14) is a consequence of the fact that $SL(2; \mathbb{Z})$ is virtually free, and (15) follows from Lemma 4.20.

If $3 \leq q \leq p - 1$, then we can prove (14) using Prop. 3.2 for $SL(q; \mathbb{Z})$. The proposition implies that there is a template for w for which every triangle is either

small or has vertices in some translate of $U(j, q - j)$ for some $1 \leq j \leq q - 1$. Call the ω -triangle corresponding to the boundary of the i th triangle w_i ; then we have $\sum_i \ell(w_i)^2 = O(\ell(w)^2)$, $\delta_\Gamma(w) \leq \sum_i \delta_\Gamma(w_i)$, and each w_i either has length $\leq c$ or has vertices in $U(j_i, q - j_i; \mathbb{Z}) \subset SL(q; \mathbb{Z}) \subset SL(p; \mathbb{Z})$. In the former case, $\delta_\Gamma(w_i) \leq \delta_\Gamma(c)$. In the latter case, we use the inductive hypotheses.

Since $j_i \leq p - 2$ and $q - j_i \leq p - 2$, we can apply Prop. 4.7 to show that there are words $p_j(w_i)$, $j = 1, 2$, in $\widehat{\Sigma}_{j_i}$ and $\widehat{\Sigma}_{q-j_i}$ such that

$$\delta_\Gamma(w_i) = O(\ell(w_i)^2) + \widehat{\delta}(p_1(w_i)) + \widehat{\delta}(p_2(w_i)).$$

By induction, the latter two terms are both $O(\ell(w_i)^2)$. Thus

$$\delta_\Gamma(w) = \sum_i O(\ell(w_i)^2) = O(\ell(w)^2).$$

The second condition, (15), follows from Lemma 4.4; if w is a word in $\widehat{\Sigma}_q$, then $\widehat{\delta}(w) = \delta_\Gamma(\lambda(w))$, and we can replace $\lambda(w)$ by a word of roughly the same length in Σ_q at cost $O(\widehat{\ell}(w)^2)$.

If $q = p$, there is an additional step to prove (14). As before, we can break w into w_i , but if $j_i = 1$ or $p - 1$, we need to use Lemma 4.19 to fill w_i .

Let w_i be an ω -triangle with vertices in $U(p - 1, 1)$. We can use Lemmas 4.4 to replace w_i by a word of comparable length in Σ_H at cost $O(\widehat{\ell}(w)^2)$. By Lemma 4.19, there are ω -triangles w'_i which are either short or have vertices in $U(j'_i, p - 1 - j'_i, 1)$ for some $1 \leq j'_i \leq p - 2$, and such that $\sum_i \ell(w'_i)^2 = O(\ell(w_i)^2)$. By Prop. 4.7 and the inductive hypothesis, $\delta_\Gamma(w'_i) = O(\ell(w'_i)^2)$, and so $\delta_\Gamma(w) = O(\ell(w)^2)$.

Finally, if w is a word in $\widehat{\Sigma}_q$, then $\widehat{\delta}(w) = \delta_\Gamma(\lambda(w))$, and $\widehat{\ell}(w) = \ell(\lambda(w))$. Consequently,

$$\widehat{\delta}(w) \leq \delta_\Gamma(\widehat{\ell}(w)) = O(\widehat{\ell}(w)^2)$$

as desired.

REFERENCES

- [BHC62] A. Borel and Harish-Chandra. Arithmetic subgroups of algebraic groups. *Ann. of Math.* (2), 75:485–535, 1962.
- [Bri02] M. R. Bridson. The geometry of the word problem. In *Invitations to geometry and topology*, volume 7 of *Oxf. Grad. Texts Math.*, pages 29–91. Oxford Univ. Press, Oxford, 2002.
- [BT02] J. Burillo and J. Taback. Equivalence of geometric and combinatorial Dehn functions. *New York J. Math.*, 8:169–179 (electronic), 2002.
- [BW07] Kai-Uwe Bux and Kevin Wortman. Finiteness properties of arithmetic groups over function fields. *Invent. Math.*, 167(2):355–378, 2007.
- [dC08] Y. de Cornulier. personal communication, 2008.
- [Din94] J. T. Ding. A proof of a conjecture of C. L. Siegel. *J. Number Theory*, 46(1):1–11, 1994.
- [Dru98] Cornelia Druțu. Remplissage dans des réseaux de \mathbf{Q} -rang 1 et dans des groupes résolubles. *Pacific J. Math.*, 185(2):269–305, 1998.
- [Dru04] C. Druțu. Filling in solvable groups and in lattices in semisimple groups. *Topology*, 43(5):983–1033, 2004.
- [ECH⁺92] D. B. A. Epstein, J. W. Cannon, D. F. Holt, S. V. F. Levy, M. S. Paterson, and W. P. Thurston. *Word processing in groups*. Jones and Bartlett Publishers, Boston, MA, 1992.
- [FF60] H. Federer and W. H. Fleming. Normal and integral currents. *Ann. of Math.* (2), 72:458–520, 1960.

- [Ger93] Steve M. Gersten. Isoperimetric and isodiametric functions of finite presentations. In *Geometric group theory, Vol. 1 (Sussex, 1991)*, volume 181 of *London Math. Soc. Lecture Note Ser.*, pages 79–96. Cambridge Univ. Press, Cambridge, 1993.
- [GHR03] S. M. Gersten, D. F. Holt, and T. R. Riley. Isoperimetric inequalities for nilpotent groups. *Geom. Funct. Anal.*, 13(4):795–814, 2003.
- [Gro87] M. Gromov. Hyperbolic groups. In *Essays in group theory*, volume 8 of *Math. Sci. Res. Inst. Publ.*, pages 75–263. Springer, New York, 1987.
- [Gro93] M. Gromov. Asymptotic invariants of infinite groups. In *Geometric group theory, Vol. 2 (Sussex, 1991)*, volume 182 of *London Math. Soc. Lecture Note Ser.*, pages 1–295. Cambridge Univ. Press, Cambridge, 1993.
- [Gro96] M. Gromov. Carnot-Carathéodory spaces seen from within. In *Sub-Riemannian geometry*, volume 144 of *Progr. Math.*, pages 79–323. Birkhäuser, Basel, 1996.
- [Ji98] L. Ji. Metric compactifications of locally symmetric spaces. *Internat. J. Math.*, 9(4):465–491, 1998.
- [JM02] L. Ji and R. MacPherson. Geometry of compactifications of locally symmetric spaces. *Ann. Inst. Fourier (Grenoble)*, 52(2):457–559, 2002.
- [Leu04a] E. Leuzinger. On polyhedral retracts and compactifications of locally symmetric spaces. *Differential Geom. Appl.*, 20(3):293–318, 2004.
- [Leu04b] E. Leuzinger. Tits geometry, arithmetic groups, and the proof of a conjecture of Siegel. *J. Lie Theory*, 14(2):317–338, 2004.
- [LMR93] A. Lubotzky, S. Mozes, and M. S. Raghunathan. Cyclic subgroups of exponential growth and metrics on discrete groups. *C. R. Acad. Sci. Paris Sér. I Math.*, 317(8):735–740, 1993.
- [LP96] E. Leuzinger and Ch. Pittet. Isoperimetric inequalities for lattices in semisimple Lie groups of rank 2. *Geom. Funct. Anal.*, 6(3):489–511, 1996.
- [LP04] E. Leuzinger and Ch. Pittet. On quadratic Dehn functions. *Math. Z.*, 248(4):725–755, 2004.
- [Mil71] J. Milnor. *Introduction to algebraic K-theory*. Princeton University Press, Princeton, N.J., 1971. Annals of Mathematics Studies, No. 72.
- [Pit97] Ch. Pittet. Isoperimetric inequalities in nilpotent groups. *J. London Math. Soc. (2)*, 55(3):588–600, 1997.
- [You] R. Young. A polynomial isoperimetric inequality for $SL(n, \mathbb{Z})$, arXiv:math.GR/0903.2495.

INSTITUT DES HAUTES ÉTUDES SCIENTIFIQUES, LE BOIS MARIE, 35 ROUTE DE CHARTRES, F-91440 BURES-SUR-YVETTE, FRANCE
E-mail address: `rjyoung@ihes.fr`